PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G01N 33/48 — A2

(11) International Publication Number: **WO 00/22430**

(43) International Publication Date: 20 April 2000 (20.04.00)

(21) International Application Number: PCT/US99/23573

(22) International Filing Date: 8 October 1999 (08.10.99)

(30) Priority Data:
60/103,794 — 9 October 1998 (09.10.98) — US
60/132,068 — 30 April 1999 (30.04.99) — US

(71) Applicant *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: FRAZER, Claire, M. [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). HICKEY, Erin [US/US]; 4569 Horton Street, Emeryville, CA 94608-2916 (US). PETERSON, Jeremy [US/US]; 4569 Horton Street, Emeryville, CA 94608-2916 (US). TETTELIN, Herve [US/US]; 4569 Horton Street, Emeryville, CA 94608-2916 (US). VENTER, J., Craig [US/US]; 4569 Horton Street, Emeryville, CA 94608-2916 (US). MASIGNANI, Vega [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). GALEOTTI, Cesira [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). MORA, Marirosa [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). RATTI, Giulio [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). SCARSELLI, Maria [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). RAPPUOLI, Rino [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). PIZZA, Mariagratia [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT).

(74) Agent: HARBIN, Alisa, A.; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

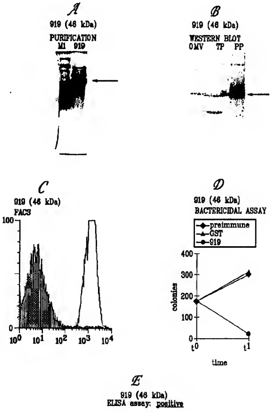
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NEISSERIA GENOMIC SEQUENCES AND METHODS OF THEIR USE

(57) Abstract

The invention provides methods of obtaining immunogenic proteins from genomic sequences including *Neisseria*, including the amino acid sequences and the corresponding nucleotide sequences, as well as the genomic sequence of *Neisseria meningitidis B*. The proteins so obtained are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NEISSERIA GENOMIC SEQUENCES AND METHODS OF THEIR USE

This application claims priority to provisional U.S. applications serial nos. 60/103,794, filed 9 October, 1998 and 60/132,068, filed 30 April, 1999, both of which are incorporated in full herein by reference.

This invention relates to methods of obtaining antigens and immunogens, the antigens and immunogens so obtained, and nucleic acids from the bacterial species: *Neisseria meningitidis*. In particular, it relates to genomic sequences from the bacterium; more particularly its "B" serogroup.

BACKGROUND

Neisseria meningitidis is a non-motile, gram negative diplococcus human pathogen. It colonizes the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. It is closely related to *N. gonorrhoea*, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N. meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks. (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N. meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for

the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants (e.g., Morbidity and Mortality weekly report, Vol. 46, No. RR-5 (1997)). This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H. influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease". In: *New Generation Vaccines*, *supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A (menA) and C (menC) (*Vaccine* 10:691-698)).

Meningococcus B (MenB) remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the MenB capsular polysaccharide is a polymer of α(2-8)-linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to MenB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (e.g., Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28).

Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (e.g., Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (e.g., EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable or at least are more antigenically conserved than other and more variable regions. Thus, those antigenic sequences that are more highly conserved are preferred sequences. Those sequences specific to *Neisseria meningitidis* or *Neisseria gonorrhoeae* that are more highly conserved are further preferred sequences. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*. The identification of sequences from the bacterium will also facilitate the production of biological probes, particularly organism-specific probes.

It is thus an object of the invention is to provide Neisserial DNA sequences which (1) encode proteins predicted and/or shown to be antigenic or immunogenic, (2) can be used as probes or amplification primers, and (3) can be analyzed by bioinformatics.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates the products of protein expression and purification of the predicted ORF 919 as cloned and expressed in *E. coli*.

Fig. 2 illustrates the products of protein expression and purification of the predicted ORF 279 as cloned and expressed in *E. coli*.

Fig. 3 illustrates the products of protein expression and purification of the predicted ORF 576-1 as cloned and expressed in *E. coli*.

Fig. 4 illustrates the products of protein expression and purification of the predicted ORF 519-1 as cloned and expressed in *E. coli*.

Fig. 5 illustrates the products of protein expression and purification of the predicted ORF 121-1 as cloned and expressed in *E. coli.*

Fig. 6 illustrates the products of protein expression and purification of the predicted ORF 128-1 as cloned and expressed in *E. coli.*

Fig. 7 illustrates the products of protein expression and purification of the predicted ORF 206 as cloned and expressed in *E. coli.*

Fig. 8 illustrates the products of protein expression and purification of the predicted ORF 287 as cloned and expressed in *E. coli.*

Fig. 9 illustrates the products of protein expression and purification of the predicted ORF 406 as cloned and expressed in *E. coli.*

Fig. 10 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 919 as cloned and expressed in *E. coli.*

Fig. 11 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 279 as cloned and expressed in *E. coli.*

Fig. 12 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 576-1 as cloned and expressed in *E. coli.*

Fig. 13 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 519-1 as cloned and expressed in *E. coli.*

Fig. 14 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 121-1 as cloned and expressed in *E. coli.*

Fig. 15 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 128-1 as cloned and expressed in *E. coli.*

Fig. 16 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 206 as cloned and expressed in *E. coli.*

Fig. 17 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 287 as cloned and expressed in *E. coli.*

Fig. 18 illustrates the hydrophilicity plot, antigenic index and AMPHI regions of the products of protein expression the predicted ORF 406 as cloned and expressed in *E. coli.*

THE INVENTION

The invention is based on the 961 nucleotide sequences from the genome of *N. meningitidis* shown as SEQ ID NOs:1-961 of Appendix C, and the full length genome of *N. meningitidis* shown as SEQ ID NO. 1068 in Appendix D. The 961 sequences in Appendix C represent substantially the whole genome of serotype B of *N. meningitidis* (>99.98%). There is partial overlap between some of the 961 contiguous sequences ("contigs") shown in the sequences in Appendix C, which overlap was used to construct the single full length sequence shown in SEQ ID NO. 1068 in Appendix D, using the TIGR Assembler [G.S. Sutton et al., *TIGR Assembler: A New Tool for Assembling Large Shotgun Sequencing Projects*, Genome Science and Technology, 1:9-19 (1995)]. Some of the nucleotides in the contigs had been previously released. (See ftp:11ftp.tigr.org/pub/data/n_meningitidis on the world-wide web or "WWW"). The coordinates of the 2508 released sequences in the present contigs are presented in Appendix A. These data include the contig number (or i.d.) as presented in the first column; the name of the sequence as found on WWW is in the second column; with the coordinates of the contigs in the third and fourth columns, respectively. The sequences of certain MenB ORFs presented in Appendix B feature in International Patent Application filed by Chiron SpA on October 9, 1998 (PCT/IB98/01665) and January 14, 1999 (PCT/IB99/00103) respectively.

In a first aspect, the invention provides nucleic acid including one or more of the *N. meningitidis* nucleotide sequences shown in SEQ ID NOs:1-961 and 1068 in Appendices C and E. It also provides nucleic acid comprising sequences having sequence identity to the nucleotide sequence disclosed herein. Depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (e.g., 60%, 70%, 80%, 90%, 95%, 99% or more). These sequences include, for instance, mutants and allelic variants. The degree of sequence identity cited herein is determined across the length of the sequence determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following parameters: gap open penalty 12, gap extension penalty 1.

The invention also provides nucleic acid including a fragment of one or more of the nucleotide sequences set out herein. The fragment should comprise at least *n* consecutive nucleotides from the sequences and, depending on the particular sequence, *n* is 10 or more

(e.g., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 75, 100 or more). Preferably, the fragment is unique to the genome of *N. meningitidis*, that is to say it is not present in the genome of another organism. More preferably, the fragment is unique to the genome of strain B of *N. meningitidis*. The invention also provides nucleic acid that hybridizes to those provided herein. Conditions for hybridizing are disclosed herein.

The invention also provides nucleic acid including sequences complementary to those described above (e.g., for antisense, for probes, or for amplification primers).

Nucleic acid according to the invention can, of course, be prepared in many ways (e.g., by chemical synthesis, from DNA libraries, from the organism itself, etc.) and can take various forms (e.g., single-stranded, double-stranded, vectors, probes, primers, etc.). The term "nucleic acid" includes DNA and RNA, and also their analogs, such as those containing modified backbones, and also peptide nucleic acid (PNA) etc.

It will be appreciated that, as SEQ ID NOs:1-961 represent the substantially complete genome of the organism, with partial overlap, references to SEQ ID NOs:1-961 include within their scope references to the complete genomic sequence, e.g., where two SEQ ID NOs overlap, the invention encompasses the single sequence which is formed by assembling the two overlapping sequences. Thus, for instance, a nucleotide sequence which bridges two SEQ ID NOs but is not present in its entirety in either SEQ ID NO is still within the scope of the invention. Additionally, such a sequence will be present in its entirety in the single full length sequence of SEQ ID NO. 1068.

The invention also provides vectors including nucleotide sequences of the invention (e.g., expression vectors, sequencing vectors, cloning vectors, etc.) and host cells transformed with such vectors.

According to a further aspect, the invention provides a protein including an amino acid sequence encoded within a *N. meningitidis* nucleotide sequence set out herein. It also provides proteins comprising sequences having sequence identity to those proteins. Depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (e.g., 60%, 70%, 80%, 90%, 95%, 99% or more). Sequence identity is determined as above disclosed. These homologous proteins include mutants and allelic variants, encoded within the *N. meningitidis* nucleotide sequence set out herein.

The invention further provides proteins including fragments of an amino acid sequence encoded within a *N. meningitidis* nucleotide sequence set out in the sequence listing. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (e.g., 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (e.g., recombinant expression, purification from cell culture, chemical synthesis, *etc.*) and in various forms (e.g. native, fusions *etc.*). They are preferably prepared in substantially isolated form (*i.e.*, substantially free from other *N. meningitidis* host cell proteins).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins of the invention. For example, the proteins can be expressed recombinantly or chemically synthesized and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question; i.e., the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The invention also provides nucleic acid encoding a protein of the invention.

In a further aspect, the invention provides a computer, a computer memory, a computer storage medium (e.g., floppy disk, fixed disk, CD-ROM, etc.), and/or a computer database containing the nucleotide sequence of nucleic acid according to the invention. Preferably, it contains one or more of the *N. meningitidis* nucleotide sequences set out herein.

This may be used in the analysis of the *N. meningitidis* nucleotide sequences set out herein. For instance, it may be used in a search to identify open reading frames (ORFs) or coding sequences within the sequences.

In a further aspect, the invention provides a method for identifying an amino acid sequence, comprising the step of searching for putative open reading frames or protein-coding sequences within a *N. meningitidis* nucleotide sequence set out herein. Similarly, the invention provides the use of a *N. meningitidis* nucleotide sequence set out herein in a search for putative open reading frames or protein-coding sequences.

Open-reading frame or protein-coding sequence analysis is generally performed on a computer using standard bioinformatic techniques. Typical algorithms or program used in the analysis include ORFFINDER (NCBI), GENMARK [Borodovsky & McIninch (1993)

Computers Chem 17:122-133], and GLIMMER [Salzberg et al. (1998) *Nucl Acids Res* 26:544-548].

A search for an open reading frame or protein-coding sequence may comprise the steps of searching a *N. meningitidis* nucleotide sequence set out herein for an initiation codon and searching the upstream sequence for an in-frame termination codon. The intervening codons represent a putative protein-coding sequence. Typically, all six possible reading frames of a sequence will be searched.

An amino acid sequence identified in this way can be expressed using any suitable system to give a protein. This protein can be used to raise antibodies which recognize epitopes within the identified amino acid sequence. These antibodies can be used to screen *N. meningitidis* to detect the presence of a protein comprising the identified amino acid sequence.

Furthermore, once an ORF or protein-coding sequence is identified, the sequence can be compared with sequence databases. Sequence analysis tools can be found at NCBI (http://www.ncbi.nlm.nih.gov) e.g., the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Suitable databases for comparison include the nonredundant GenBank, EMBL, DDBJ and PDB sequences, and the nonredundant GenBank CDS translations, PDB, SwissProt, Spupdate and PIR sequences. This comparison may give an indication of the function of a protein.

Hydrophobic domains in an amino acid sequence can be predicted using algorithms such as those based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. Hydrophobic domains represent potential transmembrane regions or hydrophobic leader sequences, which suggest that the proteins may be secreted or be surface-located. These properties are typically representative of good immunogens.

Similarly, transmembrane domains or leader sequences can be predicted using the PSORT algorithm (http://www.psort.nibb.ac.jp), and functional domains can be predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

The invention also provides nucleic acid including an open reading frame or protein-coding sequence present in a *N. meningitidis* nucleotide sequence set out herein. Furthermore, the invention provides a protein including the amino acid sequence encoded by this open reading frame or protein-coding sequence.

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means known to those skilled in the art.

The antibodies of the invention can be used in a variety of ways, e.g., for confirmation that a protein is expressed, or to confirm where a protein is expressed. Labeled antibody (e.g., fluorescent labeling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein, for instance.

According to a further aspect, the invention provides compositions including protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, as immunogenic compositions, or as diagnostic reagents.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (e.g., as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of (I) a medicament for treating or preventing infection due to Neisserial bacteria (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N. gonorrhoeae*) but are preferably *N. meningitidis*, especially strain A, strain B or strain C.

In still yet another aspect, the present invention provides for compositions including proteins, nucleic acid molecules, or antibodies. More preferable aspects of the present invention are drawn to immunogenic compositions of proteins. Further preferable aspects of the present invention contemplate pharmaceutical immunogenic compositions of proteins or vaccines and the use thereof in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria, preferably infection of MenB.

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression. A process which may further include chemical synthesis of proteins and/or chemical synthesis (at least in part) of nucleotides.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

Another aspect of the present invention provides for a process for detecting antibodies that selectably bind to antigens or polypeptides or proteins specific to any species or strain of Neisserial bacteria and preferably to strains of *N. gonorrhoeae* but more preferably to strains of *N. meningitidis*, especially strain A, strain B or strain C, more preferably MenB, where the process comprises the steps of: (a) contacting antigen or polypeptide or protein according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Methodology - Summary of standard procedures and techniques.

General

This invention provides *Neisseria meningitidis* MenB nucleotide sequences, amino acid sequences encoded therein. With these disclosed sequences, nucleic acid probe assays and expression cassettes and vectors can be produced. The proteins can also be chemically synthesized. The expression vectors can be transformed into host cells to produce proteins. The purified or isolated polypeptides can be used to produce antibodies to detect MenB proteins. Also, the host cells or extracts can be utilized for biological assays to isolate

agonists or antagonists. In addition, with these sequences one can search to identify open reading frames and identify amino acid sequences. The proteins may also be used in immunogenic compositions and as vaccine components.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature e.g., Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice,* Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C.C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference.

Expression systems

The *Neisseria* MenB nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, plant cells, baculoviruses, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A

promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation (Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual, 2nd ed.).*

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible). Depending on the promoter selected, many promotes may be inducible using known substrates, such as the use of the mouse mammary tumor virus (MMTV) promoter with the glucocorticoid responsive element (GRE) that is induced by glucocorticoid in hormone-responsive transformed cells (see for example, U.S. Patent 5,783,681).

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter (Maniatis et al. (1987) *Science 236:*1237; Alberts et al. (1989) *Molecular Biology of the Cell,* 2nd ed.). Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer (Dijkema et al (1985) *EMBO J. 4*:761) and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus (Gorman et al. (1982b) *Proc. Natl. Acad. Sci. 79*:6777) and from human cytomegalovirus (Boshart et al. (1985) *Cell 41*:521). Additionally, some enhancers are regulatable and

become active only in the presence of an inducer, such as a hormone or metal ion (Sassone-Corsi and Borelli (1986) *Trends Genet. 2*:215; Maniatis et al. (1987) Science 236:1237).

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation (Birnstiel et al. (1985) *Cell 41*:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci. 14*:105). These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 (Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*).

Usually, the above-described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of

stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 (Gluzman (1981) *Cell 23*:175) or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replication systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 (Kaufman et al. (1989) *Mol. Cell. Biol. 9*:946) and pHEBO (Shimizu et al. (1986) *Mol. Cell. Biol. 6*:1074).

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g., Hep G2), and a number of other cell lines.

ii. Plant Cellular Expression Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: U.S. 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., Nucleic Acids

Research 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reptr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only

one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of

introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an

intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) *17*:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, *42*:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli.*

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* *69*:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al.

(1988) *Gene*, *73*:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human (alpha) α-interferon, Maeda et al., (1985), *Nature 315*:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol. 8*:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, *82*:8404; mouse IL-3, (Miyajima et al., (1987) *Gene 58*:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, *7*:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) *3*:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a

restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays 4*:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. *Current Protocols in Microbiology* Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti* , *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (PCT Pub. No. WO 89/046699; Carbonell et al., (1985) *J. Virol. 56*:153; Wright (1986) *Nature 321*:718; Smith et al., (1983) *Mol. Cell. Biol. 3*:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol. 25*:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See,* e.g., Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, e.g., HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, e.g., proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the

RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (*E. coli*) (Raibaud *et al.* (1984) *Annu. Rev. Genet. 18*:173). Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) (Chang *et al.* (1977) *Nature 198*:1056), and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) (Goeddel *et al.* (1980) *Nuc. Acids Res. 8*:4057; Yelverton *et al.* (1981) *Nucl. Acids Res. 9*:731; U.S. Patent 4,738,921; EPO Publ. Nos. 036 776 and 121 775). The beta-lactamase (*bla*) promoter system (Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)), bacteriophage lambda PL (Shimatake *et al.* (1981) *Nature 292*:128) and T5 (U.S. Patent 4,689,406) promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter (U.S. Patent 4,551,433). For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor (Amann *et al.* (1983) *Gene 25*:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci. 80*:21). Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system (Studier *et al.* (1986) *J. Mol. Biol. 189*:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci. 82*:1074). In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO Publ. No. 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome

binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine *et al.* (1975) *Nature 254*:34). The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA (Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)). To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site, it is often necessary to optimize the distance between the SD sequence and the ATG of the eukaryotic gene (Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*).

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO Publ. No. 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene (Nagai *et al.* (1984) *Nature 309*:810). Fusion proteins can also be made with sequences from the *lacZ* (Jia *et al.* (1987) *Gene 60*:197), *trpE* (Allen et al. (1987) J. Biotechnol. 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol. 135*:11), and *Chey* (EPO Publ. No. 324 647) genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (e.g. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated (Miller *et al.* (1989) *Bio/Technology 7*:698).

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria (U.S. Patent 4,336,336). The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) (Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J. 3*:2437) and the *E. coli* alkaline phosphatase signal sequence (*phoA*) (Oka *et al.* (1985) *Proc. Natl. Acad. Sci. 82*:7212). As an additional example, the signal sequence of the alpha-amylase gene from various Bacillus strains can be used to secrete heterologous proteins from *B. subtilis* (Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EPO Publ. No. 244 042).

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will

generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EPO Publ. No. 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline (Davies *et al.* (1978) *Annu. Rev. Microbiol. 32*:469). Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: Bacillus subtilis (Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EPO Publ. Nos. 036 259 and 063 953; PCT Publ. No. WO 84/04541), Escherichia coli (Shimatake *et al.* (1981) *Nature 292*:128; Amann *et al.* (1985) *Gene 40*:183; Studier *et al.* (1986) *J. Mol. Biol. 189*:113; EPO Publ. Nos. 036 776, 136 829 and 136 907), Streptococcus cremoris (Powell *et al.* (1988)

Appl. Environ. Microbiol. 54:655); Streptococcus lividans (Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655), Streptomyces lividans (U.S. Patent 4,745,056).

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. (See e.g., use of *Bacillus*: Masson *et al.* (1989) *FEMS Microbiol. Lett. 60*:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EPO Publ. Nos. 036 259 and 063 953; PCT Publ. No. WO 84/04541; use of *Campylobacter*: Miller *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:856; and Wang *et al.* (1990) *J. Bacteriol. 172*:949; use of *Escherichia coli:* Cohen *et al.* (1973) *Proc. Natl. Acad. Sci. 69*:2110; Dower *et al.* (1988) *Nucleic Acids Res. 16*:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol. 53*:159; Taketo (1988) *Biochim. Biophys. Acta 949*:318; use of *Lactobacillus*: Chassy *et al.* (1987) *FEMS Microbiol. Lett. 44*:173; use of *Pseudomonas*: Fiedler *et al.* (1988) *Anal. Biochem 170*:38; use of *Staphylococcus*: Augustin *et al.* (1990) *FEMS Microbiol. Lett. 66*:203; use of *Streptococcus*: Barany *et al.* (1980) *J. Bacteriol. 144*:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun. 32*:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology 1*:412.

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS),

which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EPO Publ. No. 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO Publ. No. 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences (Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:1).

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (U.S. Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EPO Publ. No. 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, (Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA 77*:1078; Henikoff *et al.* (1981) *Nature 283*:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol. 96*:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene 11*:163; Panthier *et al.* (1980) *Curr. Genet. 2*:109;).

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the

ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, plant, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See e.g., EPO Publ. No. 196056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (e.g. ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (e.g., WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EPO Publ. No. 012 873; JPO Publ. No. 62:096,086) and the A-factor gene (U.S. Patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EPO Publ. No. 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (U.S. Patent Nos. 4,546,083 and 4,870,008; EPO Publ.

No. 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alpha factor. (See e.g., PCT Publ. No. WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 (Botstein *et al.* (1979) *Gene 8*:17-24), pCl/1 (Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA 81*:4642-4646), and YRp17 (Stinchcomb *et al.* (1982) *J. Mol. Biol. 158*:157). In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See e.g., Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome (Orr-Weaver *et al.* (1983) *Methods in Enzymol. 101*:228-245). An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for

inclusion in the vector. See Orr-Weaver *et al., supra.* One or more expression construct may integrate, possibly affecting levels of recombinant protein produced (Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:6750). The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions (Butt *et al.* (1987) *Microbiol, Rev. 51*:351).

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors and methods of introducing exogenous DNA into yeast hosts have been developed for, *inter alia*, the following yeasts: *Candida albicans* (Kurtz, *et al.* (1986) *Mol. Cell. Biol. 6*:142); *Candida maltosa* (Kunze, *et al.* (1985) *J. Basic Microbiol. 25*:141); *Hansenula polymorpha* (Gleeson, *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302); *Kluyveromyces fragilis* (Das, *et al.* (1984) *J. Bacteriol. 158*:1165); *Kluyveromyces lactis* (De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:737; Van den Berg *et al.* (1990) *Bio/Technology 8*:135); *Pichia guillerimondii* (Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141); *Pichia pastoris* (Cregg, *et al.* (1985) *Mol. Cell. Biol. 5*:3376; U.S. Patent Nos. 4,837,148 and 4,929,555); *Saccharomyces cerevisiae* (Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*:1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163); *Schizosaccharomyces*

pombe (Beach and Nurse (1981) *Nature 300*:706); and *Yarrowia lipolytica* (Davidow, *et al.* (1985) *Curr. Genet. 10*:380471 Gaillardin, *et al.* (1985) *Curr. Genet. 10*:49).

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See e.g., [Kurtz *et al.* (1986) *Mol. Cell. Biol. 6*:142; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; Candida]; [Gleeson *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol. 158*:1165; De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:1165; Van den Berg *et al.* (1990) *Bio/Technology 8*:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol. 5*:3376; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; U.S. Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*;1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163 Saccharomyces]; [Beach and Nurse (1981) *Nature 300*:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet. 10*:39; Gaillardin *et al.* (1985) *Curr. Genet. 10*:49; Yarrowia].

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be

reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as a DNA, RNA or amino acid sequence differing from but having homology with the native or disclosed sequence. Depending on the particular sequence, the degree of homology between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (e.g., 60%, 70%, 80%, 90%, 95%, 99% or more) which is calculated as described above. As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs at essentially the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions. (see, for example, U.S. Patent 5,753,235).

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanized antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying *Neisseria* MenB proteins. Antibodies elicited against the proteins of the present invention bind to antigenic polypeptides or proteins or protein fragments that are present and specifically associated with strains of *Neisseria meningitidis* MenB. In some instances, these antigens may be associated with specific strains, such as those antigens specific for the MenB strains. The antibodies of

the invention may be immobilized to a matrix and utilized in an immunoassay or on an affinity chromatography column, to enable the detection and/or separation of polypeptides, proteins or protein fragments or cells comprising such polypeptides, proteins or protein fragments. Alternatively, such polypeptides, proteins or protein fragments may be immobilized so as to detect antibodies bindably specific thereto.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (e.g., 1,000*g* for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein (*Nature* (1975) 256:495-96), or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells that express membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (e.g., hypoxanthine, aminopterin, thymidine medium,

"HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (e.g., in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I, or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Antigens, immunogens, polypeptides, proteins or protein fragments of the present invention elicit formation of specific binding partner antibodies. These antigens, immunogens, polypeptides, proteins or protein fragments of the present invention comprise immunogenic compositions of the present invention. Such immunogenic compositions may further comprise or include adjuvants, carriers, or other compositions that promote or enhance or stabilize the antigens, polypeptides, proteins or protein fragments of the present

invention. Such adjuvants and carriers will be readily apparent to those of ordinary skill in the art.

Pharmaceutical Compositions

Pharmaceutical compositions can include either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature, when given to a patient that is febrile. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgment of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal and transcutaneous applications, needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (i.e., to prevent infection) or therapeutic (i.e., to treat disease after infection).

Such vaccines comprise immunizing antigen(s) or immunogen(s), immunogenic polypeptide, protein(s) or protein fragments, or nucleic acids (e.g., ribonucleic acid or deoxyribonucleic acid), usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to

the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the immunogen or antigen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (PCT Publ. No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (e.g., gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an E. coli heat-labile toxin (LT), particularly LT-K63, LT-R72, CT-S109, PT-K9/G129; see, e.g., WO 93/13302 and WO 92/19265; and (7) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59 are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The vaccine compositions comprising immunogenic compositions (e.g., which may include the antigen, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Alternatively, vaccine compositions comprising immunogenic compositions may comprise an antigen, polypeptide, protein, protein fragment or nucleic acid in a pharmaceutically acceptable carrier.

More specifically, vaccines comprising immunogenic compositions comprise an immunologically effective amount of the immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g., nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

Typically, the vaccine compositions or immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

The immunogenic compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal and transcutaneous applications. Dosage treatment may be a single dose schedule

or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed (e.g., Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648).

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs, including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses e.g., MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (e.g., HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (i.e., there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470.

Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors comprising sequences of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in U.S. Serial No. 08/405,627, filed March 15, 1995,WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukarytic layered expression systems are also useful for expressing the nucleic acids of the invention. SeeWO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox

virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and Nature (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of

photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed to transform a host cell. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in U.S. 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of

photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in U.S. 5,149,655; use of ionizing radiation for activating transferred gene, as described in U.S. 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; inWO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprise a therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered ex vivo, to cells derived from the subject; or (3) in vitro for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, transdermally or transcutaneously, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a tumor or lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule. See WO98/20734.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in e.g., WO93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and Polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

One example are polypeptides which include, without limitation: asialoorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, Etc.

Other groups that can be included in a pharmaceutical composition include, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, *etc.*

Also, polyalkylene glycol can be included in a pharmaceutical compositions with the desired polynucleotides and/or polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethlylene glycol. In addition, mono-, di-, or polysaccarides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide) may be included in a pharmaceutical composition.

D. Lipids, and Liposomes

The desired polynucleotide or polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid or polypeptide. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N(1-2,3-dioleyloxy)propyl)-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See e.g., Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77); Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348); Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide or polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid sequences of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phopholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750.

Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443.

Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Techniologies, Inc., Stoughton, Massachusetts, USA.

Further description of lipoproteins can be found in Zuckermann et al., PCT. Appln. No. US97/14465.

F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide and/or polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples of useful polypeptides include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as ΦX174, transcriptional

factors also contain domains that bind DNA and therefore may be useful as nucleic aid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and purtrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

G. Synthetic Polycationic Agents

Synthetic polycationic agents which are useful in pharmaceutical compositions include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides or polypeptides.

Immunodiagnostic Assays

Neisseria MenB antigens, or antigenic fragments thereof, of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-*Neisseria* MenB antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to *Neisseria* MenB proteins or fragments thereof within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the

invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridization

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* (*supra*) Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8

cpm/μg. For a single-copy mammalian gene a conservative approach would start with 10 μg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/μg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$Tm = 81 + 16.6(\log_{10}Ci) + 0.4(\%(G + C)) - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch})$$

where Ci is the salt concentration (monovalent ions) and *n* is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138:267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (i.e., stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this

approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) -- some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least

30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* (*J. Am. Chem. Soc.* (1981) 103:3185), or according to Urdea *et al.* (*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461), or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For ccrtain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated e.g., backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* (e.g., see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387); analogues such as peptide nucleic acids may also be used (e.g., see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386).

One example of a nucleotide hybridization assay is described by Urdea *et al.* in international patent application WO92/02526 (see also U.S. Patent 5,124,246).

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* (*Meth. Enzymol.* (1987) 155: 335-350); US patent 4,683,195; and US patent 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelcd probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* (*supra*). mRNA, or cDNA generated from mRNA using a polymerase

enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labeled with a radioactive moiety.

EXAMPLES

The invention is based on the 961 nucleotide sequences from the genome of *N. meningitidis* set out in Appendix C, SEQ ID NOs:1-961, which together represent substantially the complete genome of serotype B of *N. meningitidis*, as well as the full length genome sequence shown in Appendix D, SEQ ID NO 1068.

It will be self-evident to the skilled person how this sequence information can be utilized according to the invention, as above described.

The standard techniques and procedures which may be employed in order to perform the invention (*e.g.* to utilize the disclosed sequences to predict polypeptides useful for vaccination or diagnostic purposes) were summarized above. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

These sequences are derived from contigs shown in Appendix C (SEQ ID NOs 1-961) and from the full length genome sequence shown in Appendix D (SEQ ID NO 1068), which were prepared during the sequencing of the genome of *N. meningitidis* (strain B). The full length sequence was assembled using the TIGR Assembler as described by G.S. Sutton et al., *TIGR Assembler: A New Tool for Assembling Large Shotgun Sequencing Projects*, Genome Science and Technology, 1:9-19 (1995) [*see also* R. D. Fleischmann, et al., Science 269, 496-512 (1995); C. M. Fraser, et al., Science 270, 397-403 (1995); C. J. Bult, et al., Science 273, 1058-73 (1996); C. M. Fraser, et. al, Nature 390, 580-586 (1997); J.-F. Tomb, et. al., Nature 388, 539-547 (1997); H. P. Klenk, et al., Nature 390, 364-70 (1997); C. M. Fraser, et al., Science 281, 375-88 (1998); M. J. Gardner, et al., Science 282, 1126-1132 (1998); K. E. Nelson, et al., Nature 399, 323-9 (1999)]. Then, using the above-described methods, putative translation products of the sequences were determined. Computer analysis of the translation products were determined based on database comparisons. Corresponding gene and protein sequences, if any, were identified in Neisseria meningitidis (Strain A) and Neisseria

gonorrhoeae. Then the proteins were expressed, purified, and characterized to assess their antigenicity and immunogenicity.

In particular, the following methods were used to express, purify, and biochemically characterize the proteins of the invention.

Chromosomal DNA Preparation

N. meningitidis strain 2996 was grown to exponential phase in 100 ml of GC medium, harvested by centrifugation, and resuspended in 5 ml buffer (20% Sucrose, 50 mM Tris-HCl, 50 mM EDTA, adjusted to pH 8.0). After 10 minutes incubation on ice, the bacteria were lysed by adding 10 ml lysis solution (50 mM NaCl, 1% Na-Sarkosyl, 50 μg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one $ChCl_3$/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4 ml buffer (10 mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*BamH*I-*Nde*I, *BamH*I-*Nhe*I, or *EcoR*I-*Nhe*I, depending on the gene's restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamH*I-*Xho*I or *EcoR*I-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

5'-end primer tail:	CGCGGATCCCATATG	(*BamH*I-*Nde*I)
	CGCGGATCCGCTAGC	(*BamH*I-*Nhe*I)

	CCGGAATTCTAGCTAGC	(*EcoR*I-*Nhe*I)
3'-end primer tail:	CCCGCTCGAG	(*Xho*I)

For some ORFs, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

5'-end primer tail:	GGAATTCCATATGGCCATGG	(*Nde*I)
5'-end primer tail:	CGGGATCC	(*BamH*I)

Other ORFs were cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. The predicted signal peptide may be included in the final product. *Nhe*I-*BamH*I restriction sites were incorporated using primers:

5'-end primer tail:	GATCAGCTAGCCATATG	(*Nhe*I)
3'-end primer tail:	CGGGATCC	(*BamH*I)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridizeed to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$T_m = 4\,(G+C) + 2\,(A+T)$ (tail excluded)

$T_m = 64.9 + 0.41\,(\%\,GC) - 600/N$ (whole primer)

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2 ml NH_4-OH, and deprotected by 5 hours incubation at 56 °C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectophotometer and the concentration was determined and adjusted to 2-10 pmol/µl.

Table 1 shows the forward and reverse primers used for each amplification. In certain cases, it might be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications are performed, the complete 5' and/or 3' sequence

may not be known for some meningococcal ORFs, although the corresponding sequences may have been identified in gonoccus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons may be changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA and CGG→CGC; GGG→GGC.

Amplification

The standard PCR protocol was as follows: 50-200 ng of genomic DNA were used as a template in the presence of 20-40 μM of each oligo, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5 mM $MgCl_2$), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimsed by the addition of 10μl DMSO or 50 μl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds 95°C	30 seconds 65-70°C	30-60 seconds 72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose

gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30μl or 50μl of either water or 10mM Tris, pH 8.5.

Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

NdeI/*Xho*I or *NheI*/*Xho*I for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion

BamHI/*Xho*I or *EcoRI*/*Xho*I for cloning into pGEX-KG and further expression of the protein as a GST N-terminus fusion.

For ORF 76, *NheI*/*BamH*I for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40 μl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 (or 50) μl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

Digestion of the cloning vectors (pET22B, pGEX-KG and pTRC-His A)

10 μg plasmid was double-digested with 50 units of each restriction enzyme in 200 μl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50 μl of 10 mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD_{260} of the sample, and adjusted to 50 μg/μl. 1 μl of plasmid was used for each cloning procedure.

Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20 µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5 µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100 µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800 µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200 µl of the supernatant. The suspension was then plated on LB ampicillin (100 mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37 °C in either 2 ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100 µg/ml ampicillin. The cells were then pelletted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30 µl. 5 µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

Cloning

Certain ORFs may be cloned into the pGEX-HIS vector using *EcoRI-PstI*, *EcoRI-SalI*, or *SalI-PstI* cloning sites. After cloning, the recombinant plasmids may be introduced in the *E.*coli host W3110.

Expression

Each ORF cloned into the expression vector may then be transformed into the strain suitable for expression of the recombinant protein product. 1 µl of each construct was used to

transform 30 μl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100 μg/ml), incubated at 37°C overnight, then diluted 1:30 in 20 ml of LB+Amp (100 μg/ml) in 100 ml flasks, making sure that the OD_{600} ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addiction of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2 mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000*g* and the pellet resuspended in PBS for further use.

GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20 ml of LB+Amp liquid colture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600 ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD_{550} 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000 rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5 ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again. The supernatant was collected and mixed with 150μl Glutatione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700*g* for 5 minutes at 4C. The resin was washed twice with 10 ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD_{280} of 0.02-0.06. The GST-fusion protein was eluted by addition of 700μl cold Glutathione elution buffer 10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD_{280} was 0.1.

21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M") (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

His-fusion soluble proteins large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD_{550} 0.6-0.8. Protein expression was induced by addition of 1 mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000 rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold 10mM imidazole buffer (300 mM NaCl, 50 mM phosphate buffer, 10 mM imidazole, pH 8). The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again. The supernatant was collected and mixed with 150µl Ni^{2+}-resin (Pharmacia) (previously washed with 10mM imidazole buffer) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10 ml cold 10mM imidazole buffer for 10 minutes, resuspended in 1ml cold 10mM imidazole buffer and loaded on a disposable column. The resin was washed at 4°C with 2ml cold 10mM imidazole buffer until the flow-through reached the $O.D_{280}$ of 0.02-0.06. The resin was washed with 2ml cold 20mM imidazole buffer (300 mM NaCl, 50 mM phosphate buffer, 20 mM imidazole, pH 8) until the flow-through reached the $O.D_{280}$ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl cold 250mM imidazole buffer (300 mM NaCl, 50 mM phosphate buffer, 250 mM imidazole, pH 8) and fractions collected until the $O.D_{280}$ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

His-fusion insoluble proteins large-scale purification.

A single colony was grown overnight at 37 °C on a LB + Amp agar plate. The bacteria were inoculated into 20 ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml fresh medium and let to grow at the

optimal temperature (37°C) to $O.D_{550}$ 0.6-0.8. Protein expression was induced by addition of 1 mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5 ml buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8). The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was stored at -20°C, while the pellets were resuspended in 2 ml guanidine buffer (6M guanidine hydrochloride, 100mM phosphate buffer, 10 mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000 rpm for 40 minutes. The supernatant was mixed with 150μl Ni^{2+}-resin (Pharmacia) (previously washed with buffer B) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700 g for 5 minutes at 4°C. The resin was washed twice with 10 ml buffer B for 10 minutes, resuspended in 1ml buffer B, and loaded on a disposable column. The resin was washed at room temperature with 2ml buffer B until the flow-through reached the OD_{280} of 0.02-0.06. The resin was washed with 2ml buffer C (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the $O.D_{280}$ of 0.02-0.06. The His-fusion protein was eluted by addition of 700μl elution buffer (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the OD_{280} was 0.1. 21μl of each fraction were loaded on a 12% SDS gel.

His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20μg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

Mice immunisations

20μg of each purified protein were used to immunise mice intraperitoneally. In the case of some ORFs, Balb-C mice were immunised with $Al(OH)_3$ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For other ORFs, CD1 mice could be immunised using the same protocol. For other ORFs, CD1 mice could be immunised using Freund's adjuvant, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for still other ORFs, CD1 mice could be immunised with Freund's adjuvant, but the immune response was measured on day 49.

ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000 rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100μl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200 μl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200 μl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN_3 in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100 μl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100 μl of substrate buffer for HRP (25 ml of citrate buffer pH5, 10 mg of O-phenildiamine and 10 μl of H_2O) were added to each well and the plates were left at room temperature for 20 minutes. 100 μl H_2SO_4 was added to each

well and OD_{490} was followed. The ELISA was considered positive when OD490 was 2.5 times the respective pre-immune sera.

FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000 rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN_3) and centrifuged for 5 minutes at 4000 rpm. Cells were resuspended in blocking buffer to reach OD_{620} of 0.07. 100μl bacterial cells were added to each well of a Costar 96 well plate. 100μl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000 rpm, the supernatant aspirated and cells washed by addition of 200μl/well of blocking buffer in each well. 100μl of R-Phicoerytrin conjugated $F(ab)_2$ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200μl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200μl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H Treshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539. Compensation values: 0.

OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10' on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000*g* for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000*g* at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in

2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000*g* for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000*g* for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30' minutes.

Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5 μg) and total cell extracts (25μg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with 1:200 mice sera diluted in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labeled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD_{620} was in between 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD_{620} of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50μl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25μl of diluted (1:100) mice sera (dilution buffer: Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25μl of the previously described bacterial suspension were added to each well. 25μl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22μl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22μl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1h were counted.

The following DNA and amino acid sequences are identified by titles of the following form: [g, m, or a] [#].[seq or pep], where "g" means a sequence from *N. gonorrhoeae*, "m" means a sequence from *N. meningitidis B*, and "a" means a sequence from *N. meningitidis A*; "#" means the number of the sequence; "seq" means a DNA sequence, and "pep" means an amino acid sequence. For example, "g001.seq" refers to an *N. gonorrohoeae* DNA sequence, number 1. The presence of the suffix "-1" or "-2" to these sequences indicates an additional sequence found for the same ORF. Further, open reading frames are identified as ORF #, where "#" means the number of the ORF, corresponding to the number of the sequence which encodes the ORF, and the ORF designations may be suffixed with ".ng" or ".a", indicating that the ORF corresponds to a *N. gonorrhoeae* sequence or a *N. meningitidis A* sequence, respectively. Computer analysis was performed for the comparisons that follow between "g", "m", and "a" peptide sequences; and therein the "pep" suffix is implied where not expressly stated.

EXAMPLE 1

The following ORFs were predicted from the contig sequences and/or the full length sequence using the methods herein described.

Localization of the ORFs

ORF:	contig:
279	gnm4.seq

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 962>:

m279.seq

```
  1  ATAACGCGGA TTTGCGGCTG CTTGATTTCA ACGGTTTTCA GGGCTTCGGC
 51  AAGTTTGTCG GCGGCGGGTT TCATCAGGCT GCAATGGGAA GGTACGGACA
101  CGGGCAGCGG CAGGGCGCGT TTGGCACCGG CTTCTTTGGC GGCAGCCATG
151  GCGCGTCCGA CGGCGGCGGC GTTGCCTGCA ATCACGATTT GTCCGGGTGA
201  GTTGAAGTTG ACGGCTTCGA CCACTTCGCT TTGGGCGGCT TCGGCACAAA
251  TGGCTTTAAC CTGCTCATCT TCCAAGCCGA GAATCGCCGC CATTGCGCCC
301  ACGCCTTGCG GTACGGCGGA CTGCATCAGT TCGGCGCGCA GGCGCACGAG
351  TTTGACCGCG TCGGCAAAAT TCAATGCGCC GGCGGCAACG AGTGCGGTGT
401  ATTCGCCGAG GCTGTGTCCG GCAACGGCGG CAGGCGTTTT GCCGCCCGCT
451  TCTAAATAG
```

This corresponds to the amino acid sequence <SEQ ID 963; ORF 279>:

m279.pep

```
  1  ITRICGCLIS TVFRASASLS AAGFIRLQWE GTDTGSGRAR LAPASLAAAM
 51  ARPTAAALPA ITICPGELKL TASTTSLWAA SAQMALTCSS SKPRIAAIAP
101  TPCGTADCIS SARRRTSLTA SAKFNAPAAT SAVYSPRLCP ATAAGVLPPA
151  SK*
```

The following partial DNA sequence was identified in *N.gonorrhoeae* <SEQ ID 964>:

g279.seq

```
  1  atgacgcgga tttgcggctg cttgatttca acggttttga gtgtttcggc
 51  aagtttgtcg gcggcgggtt tcatcaggct gcaatgggaa ggaacggata
101  ccggcagcgg cagggcgcgt ttggctccgg cttctttggc ggcagccatg
151  gtgcgtccga cggcggcggc gttgcctgca atcacgactt gtccgggcga
201  gttgaagttg acggcttcga ccacttcgcc ctgtgcggat tcggcacaaa
251  tctgcctgac ctgttcatct tccaaaccca aaatggccgc cattgcgcct
301  acgccttgcg gtacggcgga ctgcatcagt tcggcgcgca ggcggacgag
351  tttgacggca tcggcaaaat ccaatgcttc ggcggcgaca agcgcggtgt
401  attcgccgag gctgtgtccg gcaacggcgg caggcgtttt gccgcccact
451  tccaaatag
```

This corresponds to the amino acid sequence <SEQ ID 965; ORF 279.ng>:

g279.pep

```
  1  MTRICGCLIS TVLSVSASLS AAGFIRLQWE GTDTGSGRAR LAPASLAAAM
 51  VRPTAAALPA ITTCPGELKL TASTTSPCAD SAQICLTCSS SKPKMAAIAP
101  TPCGTADCIS SARRRTSLTA SAKSNASAAT SAVYSPRLCP ATAAGVLPPT
151  SK*
```

ORF 279 shows 89.5% identity over a 152 aa overlap with a predicted ORF (ORF 279.ng) from *N. gonorrhoeae:*

```
                   10        20        30        40        50        60
m279.pep   ITRICGCLISTVFRASASLSAAGFIRLQWEGTDTGSGRARLAPASLAAAMARPTAAALPA
           :||||||||||: :||||||||||||||||||||||||||||||||||||:|||||||||
g279       MTRICGCLISTVLSVSASLSAAGFIRLQWEGTDTGSGRARLAPASLAAAMVRPTAAALPA
                   10        20        30        40        50        60

                   70        80        90       100       110       120
m279.pep   ITICPGELKLTASTTSLWAASAQMALTCSSSKPRIAAIAPTPCGTADCISSARRRTSLTA
           || |||||||||||||  | |||: ||||||||::|||||||||||||||||||||||||
g279       ITTCPGELKLTASTTSPCADSAQICLTCSSSKPKMAAIAPTPCGTADCISSARRRTSLTA
                   70        80        90       100       110       120

                  130       140       150
m279.pep   SAKFNAPAATSAVYSPRLCPATAAGVLPPASKX
           ||| || ||||||||||||||||||||||:|||
g279       SAKSNASAATSAVYSPRLCPATAAGVLPPTSKX
                  130       140       150
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 966>:

```
a279.seq
      1  ATGACNCNGA TTTGCGGCTG CTTGATTTCA ACGGTTTNNA GGGCTTCGGC
     51  GAGTTTGTCG GCGGCGGGTT TCATGAGGCT GCAATGGGAA GGTACNGACA
    101  CNGGCAGCGG CAGGGCGCGT TTGGCGCCGG CTTCTTTGGC GGCAAGCATA
    151  GCGCGCTCGA CGGCGGCGGC ATTGCCTGCA ATCACGACTT GTCCGGGCGA
    201  GTTGAAGTTG ACGGCTTCAA CCACTTCATC CTGTGCGGAT TCGGCGCAAA
    251  TTTGTTTTAC CTGTTCATCT TCCAAGCCGA GAATCGCCGC CATTGCGCCC
    301  ACGCCTTGCG GTACGGCGGA CTGCATCAGT TCGGCGCGCA NGCGCACGAG
    351  TTTGACCGCG TCGGCAAAAT CCAATGCGCC GGCGGCAACN AGTGCGGTGT
    401  ATTCGCCGAN GCTGTGTCCG GCAACGGCGG CAGGCGTTTT GCCGCCCGCT
    451  TCCGAATAG
```

This corresponds to the amino acid sequence <SEQ ID 967; ORF 279.a>:

```
a279.pep
      1  MTXICGCLIS TVXRASASLS AAGFMRLQWE GTDTGSGRAR LAPASLAASI
     51  ARSTAAALPA ITTCPGELKL TASTTSSCAD SAQICFTCSS SKPRIAAIAP
    101  TPCGTADCIS SARXRTSLTA SAKSNAPAAT SAVYSPXLCP ATAAGVLPPA
    151  SE*
```

m279/a279 ORFs 279 and 279.a showed a 88.2% identity in 152 aa overlap

```
                    10        20        30        40        50        60
m279.pep    ITRICGCLISTVFRASASLSAAGFIRLQWEGTDTGSGRARLAPASLAAAMARPTAAALPA
            :| ||||||||| |||||||||||:|||||||||||||||||||||||::|| |||||||
a279        MTXICGCLISTVXRASASLSAAGFMRLQWEGTDTGSGRARLAPASLAASIARSTAAALPA
                    10        20        30        40        50        60

                    70        80        90       100       110       120
m279.pep    ITICPGELKLTASTTSLWAASAQMALTCSSSKPRIAAIAPTPCGTADCISSARRRTSLTA
            || |||||||||||||  | |||: :||||||||||||||||||||||||||| ||||||
a279        ITTCPGELKLTASTTSSCADSAQICFTCSSSKPRIAAIAPTPCGTADCISSARXRTSLTA
                    70        80        90       100       110       120

                   130       140       150
m279.pep    SAKFNAPAATSAVYSPRLCPATAAGVLPPASKX
            ||| |||||||||||| ||||||||||||||:|
a279        SAKSNAPAATSAVYSPXLCPATAAGVLPPASEX
                   130       140       150
```

519 and 519-1 gnm7.seq

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 968>:

```
m519.seq  (partial)
      1  ..TCCGTTATCG GGCGTATGGA GTTGGACAAA ACGTTTGAAG AACGCGACGA
     51    AATCAACAGT ACTGTTGTTG CGGCTTTGGA CGAGGCGGCC GGGgCTTgGG
    101    GTGTGAAGGT TTTGCGTTAT GAGATTAAAG ACTTGGTTCC GCCGCAAGAA
    151    ATCCTTCGCT CAATGCAGGC GCAAATTACT GCCGAACGCG AAAAACGCGC
    201    CCGTATCGCC GAATCCGAAG GTCGTAAAAT CGAACAAATC AACCTTGCCA
    251    GTGGTCAGCG CGAAGCCGAA ATCCAACAAT CCGAAGGCGA GGCTCAGGCT
    301    GCGGTCAATG CGTCAAATGC CGAGAAAATC GCCCGCATCA ACCGCGCCAA
    351    AGGTGAAGCG GAATCCTTGC GCCTTGTTGC CGAAGCCAAT GCCGAAGCCA
    401    TCCGTCAAAT TGCCGCCGCC CTTCAAACCC AAGGCGGTGC GGATGCGGTC
    451    AATCTGAAGA TTGCGGAACA ATACGTCGCT GCGTTCAACA ATCTTGCCAA
    501    AGAAAGCAAT ACGCTGATTA TGCCCGCCAA TGTTGCCGAC ATCGGCAGCC
    551    TGATTTCTGC CGGTATGAAA ATTATCGACA GCAGCAAAAC CGCCAAaTAA
```

This corresponds to the amino acid sequence <SEQ ID 969; ORF 519>:

```
m519.pep    (partial)
      1   ..SVIGRMELDK TFEERDEINS TVVAALDEAA GAWGVKVLRY EIKDLVPPQE
     51     ILRSMQAQIT AEREKRARIA ESEGRKIEQI NLASGQREAE IQQSEGEAQA
    101     AVNASNAEKI ARINRAKGEA ESLRLVAEAN AEAIRQIAAA LQTQGGADAV
    151     NLKIAEQYVA AFNNLAKESN TLIMPANVAD IGSLISAGMK IIDSSKTAK*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 970>:

```
g519.seq
      1   atggaatttt tcattatctt gttggcagcc gtcgccgttt tcggcttcaa
     51   atcctttgtc gtcatccccc agcaggaagt ccacgttgtc gaaaggctcg
    101   ggcgtttcca tcgcgccctg acggccggtt tgaatatttt gattcccttt
    151   atcgaccgcg tcgcctaccg ccattcgctg aaagaaatcc ctttagacgt
    201   acccagccag gtctgcatca cgcgcgataa tacgcaattg actgttgacg
    251   gcatcatcta tttccaagta accgatccca aactcgcctc atacggttcg
    301   agcaactaca ttatggcaat tacccagctt gcccaaacga cgctgcgttc
    351   cgttatcggg cgtatggagt tggacaaaac gtttgaagaa cgcgacgaaa
    401   tcaacagtac cgtcgtctcc gccctcgatg aagccgccgg ggcttggggt
    451   gtgaaagtcc tccgttacga aatcaaggat ttggttccgc cgcaagaaat
    501   ccttcgcgca atgcaggcac aaattaccgc cgaacgcgaa aaacgcgccc
    551   gtattgccga atccgaaggc cgtaaaatcg aacaaatcaa ccttgccagt
    601   ggtcagcgtg aagccgaaat ccaacaatcc gaaggcgagg ctcaggctgc
    651   ggtcaatgcg tccaatgccg agaaaatcgc ccgcatcaac cgcgccaaag
    701   gcgaagcgga atccctgcgc cttgttgccg aagccaatgc cgaagccaac
    751   cgtcaaattg ccgccgccct tcaaacccaa agcggggcgg atgcggtcaa
    801   tctgaagatt gcgggacaat acgttaccgc gttcaaaaat cttgccaaag
    851   aagacaatac gcggattaag cccgccaagg ttgccgaaat cgggaacccct
    901   aattttcggc ggcatgaaaa attttcgcca gaagcaaaaa cggccaaata
    951   a
```

This corresponds to the amino acid sequence <SEQ ID 971; ORF 519.ng>:

```
g519.pep
      1   MEFFIILLAA VAVFGFKSFV VIPQQEVHVV ERLGRFHRAL TAGLNILIPF
     51   IDRVAYRHSL KEIPLDVPSQ VCITRDNTQL TVDGIIYFQV TDPKLASYGS
    101   SNYIMAITQL AQTTLRSVIG RMELDKTFEE RDEINSTVVS ALDEAAGAWG
    151   VKVLRYEIKD LVPPQEILRA MQAQITAERE KRARIAESEG RKIEQINLAS
    201   GQREAEIQQS EGEAQAAVNA SNAEKIARIN RAKGEAESLR LVAEANAEAN
    251   RQIAAALQTQ SGADAVNLKI AGQYVTAFKN LAKEDNTRIK PAKVAEIGNP
    301   NFRRHEKFSP EAKTAK*
```

ORF 519 shows 87.5% identity over a 200 aa overlap with a predicted ORF (ORF 519.ng) from *N. gonorrhoeae:*

```
m519/g519

                                                  10        20        30
m519.pep                              SVIGRMELDKTFEERDEINSTVVAALDEAA
                                      |||||||||||||||||||||||:||||||
g519          YFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIGRMELDKTFEERDEINSTVVSALDEAA
                90       100       110       120       130       140

                      40        50        60        70        80        90
m519.pep      GAWGVKVLRYEIKDLVPPQEILRSMQAQITAEREKRARIAESEGRKIEQINLASGQREAE
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
g519          GAWGVKVLRYEIKDLVPPQEILRAMQAQITAEREKRARIAESEGRKIEQINLASGQREAE
               150       160       170       180       190       200

                     100       110       120       130       140       150
m519.pep      IQQSEGEAQAAVNASNAEKIARINRAKGEAESLRLVAEANAEAIRQIAAALQTQGGADAV
```

```
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||:|||||
g519        IQQSEGEAQAAVNASNAEKIARINRAKGEAESLRLVAEANAEANRQIAAALQTQSGADAV
             210       220       230       240       250       260

                   160       170       180        190       200
m519.pep    NLKIAEQYVAAFNNLAKESNTLIMPANVADIGSL-ISAGMKIIDSSKTAK
            ||||| |||:||:|||||:|| | ||:||:||:  :     |:   :||||
g519        NLKIAGQYVTAFKNLAKEDNTRIKPAKVAEIGNPNFRRHEKFSPEAKTAK
             270       280       290       300       310
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 972>:

```
a519.seq
      1  ATGGAATTTT TCATTATCTT GCTGGCAGCC GTCGTTGTTT TCGGCTTCAA
     51  ATCCTTTGTT GTCATCCCAC AGCAGGAAGT CCACGTTGTC GAAAGGCTCG
    101  GGCGTTTCCA TCGCGCCCTG ACGGCCGGTT TGAATATTTT GATTCCCTTT
    151  ATCGACCGCG TCGCCTACCG CCATTCGCTG AAAGAAATCC CTTTAGACGT
    201  ACCCAGCCAG GTCTGCATCA CGCGCGACAA TACGCAGCTG ACTGTTGACG
    251  GTATCATCTA TTTCCAAGTA ACCGACCCCA AACTCGCCTC ATACGGTTCG
    301  AGCAACTACA TTATGGCGAT TACCCAGCTT GCCCAAACGA CGCTGCGTTC
    351  CGTTATCGGG CGTATGGAAT TGGACAAAAC GTTTGAAGAA CGCGACGAAA
    401  TCAACAGCAC CGTCGTCTCC GCCCTCGATG AAGCCGCCGG AGCTTGGGGT
    451  GTGAAGGTTT TGCGTTATGA GATTAAAGAC TTGGTTCCGC CGCAAGAAAT
    501  CCTTCGCTCA ATGCAGGCGC AAATTACTGC TGAACGCGAA AAACGCGCCC
    551  GTATCGCCGA ATCCGAAGGT CGTAAAATCG AACAAATCAA CCTTGCCAGT
    601  GGTCAGCGCG AAGCCGAAAT CCAACAATCC GAAGGCGAGG CTCAGGCTGC
    651  GGTCAATGCG TCAAATGCCG AGAAAATCGC CCGCATCAAC CGCGCCAAAG
    701  GTGAAGCGGA ATCCTTGCGC CTTGTTGCCG AAGCCAATGC CGAAGCCATC
    751  CGTCAAATTG CCGCCGCCCT TCAAACCCAA GGCGGTGCGG ATGCGGTCAA
    801  TCTGAAGATT GCGGAACAAT ACGTCGCCGC GTTCAACAAT CTTGCCAAAG
    851  AAAGCAATAC GCTGATTATG CCCGCCAATG TTGCCGACAT CGGCAGCCTG
    901  ATTTCTGCCG GTATGAAAAT TATCGACAGC AGCAAAACCG CCAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 973; ORF 519.a>:

```
a519.pep
      1  MEFFIILLAA VVVFGFKSFV VIPQQEVHVV ERLGRFHRAL TAGLNILIPF
     51  IDRVAYRHSL KEIPLDVPSQ VCITRDNTQL TVDGIIYFQV TDPKLASYGS
    101  SNYIMAITQL AQTTLRSVIG RMELDKTFEE RDEINSTVVS ALDEAAGAWG
    151  VKVLRYEIKD LVPPQEILRS MQAQITAERE KRARIAESEG RKIEQINLAS
    201  GQREAEIQQS EGEAQAAVNA SNAEKIARIN RAKGEAESLR LVAEANAEAI
    251  RQIAAALQTQ GGADAVNLKI AEQYVAAFNN LAKESNTLIM PANVADIGSL
    301  ISAGMKIIDS SKTAK*
```

m519/a519 **ORFs 519 and 519.a showed a** 99.5% identity in 199 aa overlap

```
                                                 10        20        30
m519.pep                                 SVIGRMELDKTFEERDEINSTVVAALDEAA
                                         ||||||||||||||||||||||||:||||||
a519        YFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIGRMELDKTFEERDEINSTVVSALDEAA
              90       100       110       120       130       140

                   40        50        60        70        80        90
m519.pep    GAWGVKVLRYEIKDLVPPQEILRSMQAQITAEREKRARIAESEGRKIEQINLASGQREAE
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a519        GAWGVKVLRYEIKDLVPPQEILRSMQAQITAEREKRARIAESEGRKIEQINLASGQREAE
             150       160       170       180       190       200

                  100       110       120       130       140       150
m519.pep    IQQSEGEAQAAVNASNAEKIARINRAKGEAESLRLVAEANAEAIRQIAAALQTQGGADAV
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a519        IQQSEGEAQAAVNASNAEKIARINRAKGEAESLRLVAEANAEAIRQIAAALQTQGGADAV
             210       220       230       240       250       260
```

```
                 160        170        180        190        200
m519.pep   NLKIAEQYVAAFNNLAKESNTLIMPANVADIGSLISAGMKIIDSSKTAKX
           ||||||||||||||||||||||||||||||||||||||||||||||||||
a519       NLKIAEQYVAAFNNLAKESNTLIMPANVADIGSLISAGMKIIDSSKTAKX
            270        280        290        300        310
```

Further work revealed the following DNA sequence identified in *N. meningitidis* <SEQ ID 974>:

```
m519-1.seq
      1  ATGGAATTTT TCATTATCTT GTTGGTAGCC GTCGCCGTTT TCGGTTTCAA
     51  ATCCTTTGTT GTCATCCCAC AACAGGAAGT CCACGTTGTC GAAAGGCTGG
    101  GGCGTTTCCA TCGCGCCCTG ACGGcCGGTT TGAATATTTT GATTCCCTTT
    151  ATCGACCGCG TCGCCTACCG CCATTCGCTG AAAGAAATCC CTTTAGACGT
    201  ACCCAGCCAG GTCTGCATCA CGCGCGACAA TACGCAGCTG ACTGTTGACG
    251  GCATCATCTA TTTCCAAGTA ACCGACCCCA AACTCGCCTC ATACGGTTCG
    301  AGCAACTACA TTATGGCGAT TACCCAGCTT GCCCAAACGA CGCTGCGTTC
    351  CGTTATCGGG CGTATGGAGT TGGACAAAAC GTTTGAAGAA CGCGACGAAA
    401  TCAACAGTAC TGTTGTTGCG GCTTTGGACG AGGCGGCCGG GGCTTGGGGT
    451  GTGAAGGTTT TGCGTTATGA GATTAAAGAC TTGGTTCCGC CGCAAGAAAT
    501  CCTTCGCTCA ATGCAGGCGC AAATTACTGC CGAACGCGAA AAACGCGCCC
    551  GTATCGCCGA ATCCGAAGGT CGTAAAATCG AACAAATCAA CCTTGCCAGT
    601  GGTCAGCGCG AAGCCGAAAT CCAACAATCC GAAGGCGAGG CTCAGGCTGC
    651  GGTCAATGCG TCAAATGCCG AGAAAATCGC CCGCATCAAC CGCGCCAAAG
    701  GTGAAGCGGA ATCCTTGCGC CTTGTTGCCG AAGCCAATGC CGAAGCCATC
    751  CGTCAAATTG CCGCCGCCCT TCAAACCCAA GGCGGTGCGG ATGCGGTCAA
    801  TCTGAAGATT GCGGAACAAT ACGTCGCTGC GTTCAACAAT CTTGCCAAAG
    851  AAAGCAATAC GCTGATTATG CCCGCCAATG TTGCCGACAT CGGCAGCCTG
    901  ATTTCTGCCG GTATGAAAAT TATCGACAGC AGCAAAACCG CCAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 975; ORF 519-1>:

```
m519-1.
      1  MEFFIILLVA VAVFGFKSFV VIPQQEVHVV ERLGRFHRAL TAGLNILIPF
     51  IDRVAYRHSL KEIPLDVPSQ VCITRDNTQL TVDGIIYFQV TDPKLASYGS
    101  SNYIMAITQL AQTTLRSVIG RMELDKTFEE RDEINSTVVA ALDEAAGAWG
    151  VKVLRYEIKD LVPPQEILRS MQAQITAERE KRARIAESEG RKIEQINLAS
    201  GQREAEIQQS EGEAQAAVNA SNAEKIARIN RAKGEAESLR LVAEANAEAI
    251  RQIAAALQTQ GGADAVNLKI AEQYVAAFNN LAKESNTLIM PANVADIGSL
    301  ISAGMKIIDS SKTAK*
```

The following DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 976>:

```
g519-1.seq
      1  ATGGAATTTT TCATTATCTT GTTGGCAGCC GTCGCCGTTT TCGGCTTCAA
     51  ATCCTTTGTC GTCATCCCCC AGCAGGAAGT CCACGTTGTC GAAAGGCTCG
    101  GGCGTTTCCA TCGCGCCCTG ACGGCCGGTT TGAATATTTT GATTCCCTTT
    151  ATCGACCGCG TCGCCTACCG CCATTCGCTG AAAGAAATCC CTTTAGACGT
    201  ACCCAGCCAG GTCTGCATCA CGCGCGATAA TACGCAATTG ACTGTTGACG
    251  GCATCATCTA TTTCCAAGTA ACCGATCCCA AACTCGCCTC ATACGGTTCG
    301  AGCAACTACA TTATGGCAAT TACCCAGCTT GCCCAAACGA CGCTGCGTTC
    351  CGTTATCGGG CGTATGGAGT TGGACAAAAC GTTTGAAGAA CGCGACGAAA
    401  TCAACAGTAC CGTCGTCTCC GCCCTCGATG AAGCCGCCGG GGCTTGGGGT
    451  GTGAAAGTCC TCCGTTACGA AATCAAGGAT TTGGTTCCGC CGCAAGAAAT
    501  CCTTCGCGCA ATGCAGGCAC AAATTACCGC CGAACGCGAA AAACGCGCCC
    551  GTATTGCCGA ATCCGAAGGC CGTAAAATCG AACAAATCAA CCTTGCCAGT
    601  GGTCAGCGTG AAGCCGAAAT CCAACAATCC GAAGGCGAGG CTCAGGCTGC
    651  GGTCAATGCG TCCAATGCCG AGAAAATCGC CCGCATCAAC CGCGCCAAAG
    701  GCGAAGCGGA ATCCCTGCGC CTTGTTGCCG AAGCCAATGC CGAAGCCATC
    751  CGTCAAATTG CCGCCGCCCT TCAAACCCAA GGCGGGGCGG ATGCGGTCAA
    801  TCTGAAGATT GCGGAACAAT ACGTAGCCGC GTTCAACAAT CTTGCCAAAG
```

```
851  AAAGCAATAC GCTGATTATG CCCGCCAATG TTGCCGACAT CGGCAGCCTG
901  ATTTCTGCCG GCATGAAAAT TATCGACAGC AGCAAAACCG CCAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 977; ORF 519-1.ng>:

```
g519-1.pep
  1  MEFFIILLAA VAVFGFKSFV VIPQQEVHVV ERLGRFHRAL TAGLNILIPF
 51  IDRVAYRHSL KEIPLDVPSQ VCITRDNTQL TVDGIIYFQV TDPKLASYGS
101  SNYIMAITQL AQTTLRSVIG RMELDKTFEE RDEINSTVVS ALDEAAGAWG
151  VKVLRYEIKD LVPPQEILRA MQAQITAERE KRARIAESEG RKIEQINLAS
201  GQREAEIQQS EGEAQAAVNA SNAEKIARIN RAKGEAESLR LVAEANAEAI
251  RQIAAALQTQ GGADAVNLKI AEQYVAAFNN LAKESNTLIM PANVADIGSL
301  ISAGMKIIDS SKTAK*
```

m519-1/g519-1 ORFs 519-1 and 519-1.ng showed a 99.0% identity in 315 aa overlap

```
                     10        20        30        40        50        60
g519-1.pep   MEFFIILLAAVAVFGFKSFVVIPQQEVHVVERLGRFHRALTAGLNILIPFIDRVAYRHSL
             ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       MEFFIILLVAVAVFGFKSFVVIPQQEVHVVERLGRFHRALTAGLNILIPFIDRVAYRHSL
                     10        20        30        40        50        60

                     70        80        90       100       110       120
g519-1.pep   KEIPLDVPSQVCITRDNTQLTVDGIIYFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       KEIPLDVPSQVCITRDNTQLTVDGIIYFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIG
                     70        80        90       100       110       120

                    130       140       150       160       170       180
g519-1.pep   RMELDKTFEERDEINSTVVSALDEAAGAWGVKVLRYEIKDLVPPQEILRAMQAQITAERE
             |||||||||||||||||||:|||||||||||||||||||||||||||||:||||||||||
m519-1       RMELDKTFEERDEINSTVVAALDEAAGAWGVKVLRYEIKDLVPPQEILRSMQAQITAERE
                    130       140       150       160       170       180

                    190       200       210       220       230       240
g519-1.pep   KRARIAESEGRKIEQINLASGQREAEIQQSEGEAQAAVNASNAEKIARINRAKGEAESLR
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       KRARIAESEGRKIEQINLASGQREAEIQQSEGEAQAAVNASNAEKIARINRAKGEAESLR
                    190       200       210       220       230       240

                    250       260       270       280       290       300
g519-1.pep   LVAEANAEAIRQIAAALQTQGGADAVNLKIAEQYVAAFNNLAKESNTLIMPANVADIGSL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       LVAEANAEAIRQIAAALQTQGGADAVNLKIAEQYVAAFNNLAKESNTLIMPANVADIGSL
                    250       260       270       280       290       300

                    310
g519-1.pep   ISAGMKIIDSSKTAKX
             ||||||||||||||||
m519-1       ISAGMKIIDSSKTAKX
                    310
```

The following DNA sequence was identified in *N. meningitidis* <SEQ ID 978>:

```
a519-1.seq
  1  ATGGAATTTT TCATTATCTT GCTGGCAGCC GTCGTTGTTT TCGGCTTCAA
 51  ATCCTTTGTT GTCATCCCAC AGCAGGAAGT CCACGTTGTC GAAAGGCTCG
101  GGCGTTTCCA TCGCGCCCTG ACGGCCGGTT TGAATATTTT GATTCCCTTT
151  ATCGACCGCG TCGCCTACCG CCATTCGCTG AAAGAAATCC CTTTAGACGT
201  ACCCAGCCAG GTCTGCATCA CGCGCGACAA TACGCAGCTG ACTGTTGACG
251  GTATCATCTA TTTCCAAGTA ACCGACCCCA AACTCGCCTC ATACGGTTCG
301  AGCAACTACA TTATGGCGAT TACCCAGCTT GCCCAAACGA CGCTGCGTTC
```

```
351  CGTTATCGGG CGTATGGAAT TGGACAAAAC GTTTGAAGAA CGCGACGAAA
401  TCAACAGCAC CGTCGTCTCC GCCCTCGATG AAGCCGCCGG AGCTTGGGGT
451  GTGAAGGTTT TGCGTTATGA GATTAAAGAC TTGGTTCCGC CGCAAGAAAT
501  CCTTCGCTCA ATGCAGGCGC AAATTACTGC TGAACGCGAA AAACGCGCCC
551  GTATCGCCGA ATCCGAAGGT CGTAAAATCG AACAAATCAA CCTTGCCAGT
601  GGTCAGCGCG AAGCCGAAAT CCAACAATCC GAAGGCGAGG CTCAGGCTGC
651  GGTCAATGCG TCAAATGCCG AGAAAATCGC CCGCATCAAC CGCGCCAAAG
701  GTGAAGCGGA ATCCTTGCGC CTTGTTGCCG AAGCCAATGC CGAAGCCATC
751  CGTCAAATTG CCGCCGCCCT TCAAACCCAA GGCGGTGCGG ATGCGGTCAA
801  TCTGAAGATT GCGGAACAAT ACGTCGCCGC GTTCAACAAT CTTGCCAAAG
851  AAAGCAATAC GCTGATTATG CCCGCCAATG TTGCCGACAT CGGCAGCCTG
901  ATTTCTGCCG GTATGAAAAT TATCGACAGC AGCAAAACCG CCAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 979; ORF 519-1.a>:

```
a519-1.pep.
      1  MEFFIILLAA VVVFGFKSFV VIPQQEVHVV ERLGRFHRAL TAGLNILIPF
     51  IDRVAYRHSL KEIPLDVPSQ VCITRDNTQL TVDGIIYFQV TDPKLASYGS
    101  SNYIMAITQL AQTTLRSVIG RMELDKTFEE RDEINSTVVS ALDEAAGAWG
    151  VKVLRYEIKD LVPPQEILRS MQAQITAERE KRARIAESEG RKIEQINLAS
    201  GQREAEIQQS EGEAQAAVNA SNAEKIARIN RAKGEAESLR LVAEANAEAI
    251  RQIAAALQTQ GGADAVNLKI AEQYVAAFNN LAKESNTLIM PANVADIGSL
    301  ISAGMKIIDS SKTAK*
```

m519-1/a519-1 ORFs 519-1 and 519-1.a showed a 99.0% identity in 315 aa overlap

```
                     10        20        30        40        50        60
a519-1.pep   MEFFIILLAAVVVFGFKSFVVIPQQEVHVVERLGRFHRALTAGLNILIPFIDRVAYRHSL
             ||||||||:||:||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       MEFFIILLVAVAVFGFKSFVVIPQQEVHVVERLGRFHRALTAGLNILIPFIDRVAYRHSL
                     10        20        30        40        50        60

                     70        80        90       100       110       120
a519-1.pep   KEIPLDVPSQVCITRDNTQLTVDGIIYFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       KEIPLDVPSQVCITRDNTQLTVDGIIYFQVTDPKLASYGSSNYIMAITQLAQTTLRSVIG
                     70        80        90       100       110       120

                    130       140       150       160       170       180
a519-1.pep   RMELDKTFEERDEINSTVVSALDEAAGAWGVKVLRYEIKDLVPPQEILRSMQAQITAERE
             |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
m519-1       RMELDKTFEERDEINSTVVAALDEAAGAWGVKVLRYEIKDLVPPQEILRSMQAQITAERE
                    130       140       150       160       170       180

                    190       200       210       220       230       240
a519-1.pep   KRARIAESEGRKIEQINLASGQREAEIQQSEGEAQAAVNASNAEKIARINRAKGEAESLR
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       KRARIAESEGRKIEQINLASGQREAEIQQSEGEAQAAVNASNAEKIARINRAKGEAESLR
                    190       200       210       220       230       240

                    250       260       270       280       290       300
a519-1.pep   LVAEANAEAIRQIAAALQTQGGADAVNLKIAEQYVAAFNNLAKESNTLIMPANVADIGSL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m519-1       LVAEANAEAIRQIAAALQTQGGADAVNLKIAEQYVAAFNNLAKESNTLIMPANVADIGSL
                    250       260       270       280       290       300

                    310
a519-1.pep   ISAGMKIIDSSKTAKX
             ||||||||||||||||
m519-1       ISAGMKIIDSSKTAKX
                    310
```

576 and 576-1 gnm22.seq

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 980>:

```
m576.seq.. (partial)
      1  ..ATGCAGCAGG CAAGCTATGC GATGGGCGTG GACATCGGAC GCTCCCTGAA
     51    GCAAATGAAG GAACAGGGCG CGGAAATCGA TTTGAAAGTC TTTACCGAAG
    101    CCATGCAGGC AGTGTATGAC GGCAAAGAAA TCAAAATGAC CGAAGAGCAG
    151    GCTCAGGAAG TCATGATGAA ATTCCTTCAG GAACAACAGG CTAAAGCCGT
    201    AGAAAAACAC AAGGCGGACG CGAAGGCCAA TAAAGAAAAA GGCGAAGCCT
    251    TTCTGAAAGA AAATGCCGCC AAAGACGGCG TGAAGACCAC TGCTTCCGGC
    301    CTGCAATACA AAATCACCAA ACAGGGCGAA GGCAAACAGC CGACCAAAGA
    351    CGACATCGTT ACCGTGGAAT ACGAAGGCCG CCTGATTGAC GGTACGGTAT
    401    TCGACAGCAG CAAAGCCAAC GGCGGCCCGG TCACCTTCCC TTTGAGCCAA
    451    GTGATTCCGG GTTGGACCGA AGgCGTACAG CTTCTGAAAG AAGGCGGCGA
    501    AGCCACGTTC TACATCCCGT CCAACCTTGC CTACCGCGAA CAGGGTGCGG
    551    GCGACAAAAT CGGTCCGAAC GCCACTTTGG TATTTGATGT GAAACTGGTC
    601    AAAATCGGCG CACCCGAAAA CGCGCCCGCC AAGCAGCCGG CTCAAGTCGA
    651    CATCAAAAAA GTAAATTAA
```

This corresponds to the amino acid sequence <SEQ ID 981; ORF 576>:

```
m576.pep.. (partial)
      1  ..MQQASYAMGV DIGRSLKQMK EQGAEIDLKV FTEAMQAVYD GKEIKMTEEQ
     51    AQEVMMKFLQ EQQAKAVEKH KADAKANKEK GEAFLKENAA KDGVKTTASG
    101    LQYKITKQGE GKQPTKDDIV TVEYEGRLID GTVFDSSKAN GGPVTFPLSQ
    151    VIPGWTEGVQ LLKEGGEATF YIPSNLAYRE QGAGDKIGPN ATLVFDVKLV
    201    KIGAPENAPA KQPAQVDIKK VN*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 982>:

```
g576.seq..(partial)
      1  ..atgggcgtgg acatcggacg ctccctgaaa caaatgaagg aacagggcgc
     51    ggaaatcgat ttgaaagtct ttaccgatgc catgcaggca gtgtatgacg
    101    gcaaagaaat caaaatgacc gaagagcagg cccaggaagt gatgatgaaa
    151    ttcctgcagg agcagcaggc taaagccgta gaaaaacaca aggcggatgc
    201    gaaggccaac aaagaaaaag gcgaagcctt cctgaaggaa aatgccgccg
    251    aagacggcgt gaagaccact gcttccggtc tgcagtacaa aatcaccaaa
    301    cagggtgaag gcaaacagcc gacaaaagac gacatcgtta ccgtggaata
    351    cgaaggccgc ctgattgacg gtaccgtatt cgacagcagc aaagccaacg
    401    gcggcccggc caccttccct ttgagccaag tgattccggg ttggaccgaa
    451    ggcgtacggc ttctgaaaga aggcggcgaa gccacgttct acatcccgtc
    501    caaccttgcc taccgcgaac agggtgcggg cgaaaaaatc ggtccgaacg
    551    ccactttggt atttgacgtg aaactggtca aaatcggcgc acccgaaaac
    601    gcgcccgcca agcagccgga tcaagtcgac atcaaaaaag taaattaa
```

This corresponds to the amino acid sequence <SEQ ID 983; ORF 576.ng>:

```
g576.pep..(partial)
      1  ..MGVDIGRSLK QMKEQGAEID LKVFTDAMQA VYDGKEIKMT EEQAQEVMMK
     51    FLQEQQAKAV EKHKADAKAN KEKGEAFLKE NAAEDGVKTT ASGLQYKITK
    101    QGEGKQPTKD DIVTVEYEGR LIDGTVFDSS KANGGPATFP LSQVIPGWTE
    151    GVRLLKEGGE ATFYIPSNLA YREQGAGEKI GPNATLVFDV KLVKIGAPEN
    201    APAKQPDQVD IKKVN*
```

Computer analysis of this amino acid sequence gave the following results:
<u>Homology with a predicted ORF from *N. gonorrhoeae*</u>

m576/g576 97.2% identity in 215 aa overlap

```
                         10        20        30        40        50        60
m576.pep         MQQASYAMGVDIGRSLKQMKEQGAEIDLKVFTEAMQAVYDGKEIKMTEEQAQEVMMKFLQ
                         ||||||||||||||||||||||||||:||||||||||||||||||||||||||
g576                     MGVDIGRSLKQMKEQGAEIDLKVFTDAMQAVYDGKEIKMTEEQAQEVMMKFLQ
                                 10        20        30        40        50

                         70        80        90       100       110       120
m576.pep         EQQAKAVEKHKADAKANKEKGEAFLKENAAKDGVKTTASGLQYKITKQGEGKQPTKDDIV
                 ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
g576             EQQAKAVEKHKADAKANKEKGEAFLKENAAEDGVKTTASGLQYKITKQGEGKQPTKDDIV
                      60        70        80        90       100       110

                        130       140       150       160       170       180
m576.pep         TVEYEGRLIDGTVFDSSKANGGPVTFPLSQVIPGWTEGVQLLKEGGEATFYIPSNLAYRE
                 ||||||||||||||||||||||:||||||||||||||||:||||||||||||||||||||
g576             TVEYEGRLIDGTVFDSSKANGGPATFPLSQVIPGWTEGVRLLKEGGEATFYIPSNLAYRE
                      120       130       140       150       160       170

                        190       200       210       220
m576.pep         QGAGDKIGPNATLVFDVKLVKIGAPENAPAKQPAQVDIKKVNX
                 ||||:||||||||||||||||||||||||||||| |||||||||
g576             QGAGEKIGPNATLVFDVKLVKIGAPENAPAKQPDQVDIKKVNX
                      180       190       200       210
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 984>:

```
a576.seq
      1  ATGAACACCA TTTTCAAAAT CAGCGCACTG ACCCTTTCCG CCGCTTTGGC
     51  ACTTTCCGCC TGCGGCAAAA AAGAAGCCGC CCCCGCATCT GCATCCGAAC
    101  CTGCCGCCGC TTCTTCCGCG CAGGGCGACA CCTCTTCGAT CGGCAGCACG
    151  ATGCAGCAGG CAAGCTATGC GATGGGCGTG GACATCGGAC GCTCCCTGAA
    201  GCAAATGAAG GAACAGGGCG CGGAAATCGA TTTGAAAGTC TTTACCGAAG
    251  CCATGCAGGC AGTGTATGAC GGCAAAGAAA TCAAAATGAC CGAAGAGCAG
    301  GCTCAGGAAG TCATGATGAA ATTCCTTCAG GAACAACAGG CTAAAGCCGT
    351  AGAAAAACAC AAGGCGGACG CGAAGGCCAA TAAAGAAAAA GGCGAAGCCT
    401  TTCTGAAAGA AAATGCCGCC AAAGACGGCG TGAAGACCAC TGCTTCCGGC
    451  CTGCAATACA AAATCACCAA ACAGGGCGAA GGCAAACAGC CGACCAAAGA
    501  CGACATCGTT ACCGTGGAAT ACGAAGGCCG CCTGATTGAC GGTACGGTAT
    551  TCGACAGCAG CAAAGCCAAC GGCGGCCCGG TCACCTTCCC TTTGAGCCAA
    601  GTGATTCTGG GTTGGACCGA AGGCGTACAG CTTCTGAAAG AAGGCGGCGA
    651  AGCCACGTTC TACATCCCGT CCAACCTTGC CTACCGCGAA CAGGGTGCGG
    701  GCGACAAAAT CGGCCCGAAC GCCACTTTGG TATTTGATGT GAAACTGGTC
    751  AAAATCGGCG CACCCGAAAA CGCGCCCGCC AAGCAGCCGG CTCAAGTCGA
    801  CATCAAAAAA GTAAATTAA
```

This corresponds to the amino acid sequence <SEQ ID 985; ORF 576.a>:

```
a576.pep
      1  MNTIFKISAL TLSAALALSA CGKKEAAPAS ASEPAAASSA QGDTSSIGST
     51  MQQASYAMGV DIGRSLKQMK EQGAEIDLKV FTEAMQAVYD GKEIKMTEEQ
    101  AQEVMMKFLQ EQQAKAVEKH KADAKANKEK GEAFLKENAA KDGVKTTASG
    151  LQYKITKQGE GKQPTKDDIV TVEYEGRLID GTVFDSSKAN GGPVTFPLSQ
    201  VILGWTEGVQ LLKEGGEATF YIPSNLAYRE QGAGDKIGPN ATLVFDVKLV
    251  KIGAPENAPA KQPAQVDIKK VN*
```

```
m576/a576     ORFs 576 and 576.a showed a 99.5% identity in 222 aa overlap

                                                   10        20        30
m576.pep                                   MQQASYAMGVDIGRSLKQMKEQGAEIDLKV
                                           ||||||||||||||||||||||||||||||
a576             CGKKEAAPASASEPAAASSAQGDTSSIGSTMQQASYAMGVDIGRSLKQMKEQGAEIDLKV
                         30        40        50        60        70        80
```

```
                  40        50        60        70        80        90
m576.pep   FTEAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKHKADAKANKEKGEAFLKENAA
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a576       FTEAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKHKADAKANKEKGEAFLKENAA
                  90       100       110       120       130       140

                 100       110       120       130       140       150
m576.pep   KDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLIDGTVFDSSKANGGPVTFPLSQ
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a576       KDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLIDGTVFDSSKANGGPVTFPLSQ
                 150       160       170       180       190       200

                 160       170       180       190       200       210
m576.pep   VIPGWTEGVQLLKEGGEATFYIPSNLAYREQGAGDKIGPNATLVFDVKLVKIGAPENAPA
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a576       VILGWTEGVQLLKEGGEATFYIPSNLAYREQGAGDKIGPNATLVFDVKLVKIGAPENAPA
                 210       220       230       240       250       260

                 220
m576.pep   KQPAQVDIKKVNX
           |||||||||||||
a576       KQPAQVDIKKVNX
                 270
```

Further work revealed the following DNA sequence identified in *N. meningitidis* <SEQ ID 986>:

```
m576-1.seq
      1  ATGAACACCA TTTTCAAAAT CAGCGCACTG ACCCTTTCCG CCGCTTTGGC
     51  ACTTTCCGCC TGCGGCAAAA AAGAAGCCGC CCCCGCATCT GCATCCGAAC
    101  CTGCCGCCGC TTCTTCCGCG CAGGGCGACA CCTCTTCGAT CGGCAGCACG
    151  ATGCAGCAGG CAAGCTATGC GATGGGCGTG GACATCGGAC GCTCCCTGAA
    201  GCAAATGAAG GAACAGGGCG CGGAAATCGA TTTGAAAGTC TTTACCGAAG
    251  CCATGCAGGC AGTGTATGAC GGCAAAGAAA TCAAAATGAC CGAAGAGCAG
    301  GCTCAGGAAG TCATGATGAA ATTCCTTCAG GAACAACAGG CTAAAGCCGT
    351  AGAAAAACAC AAGGCGGACG CGAAGGCCAA TAAAGAAAAA GGCGAAGCCT
    401  TTCTGAAAGA AAATGCCGCC AAAGACGGCG TGAAGACCAC TGCTTCCGGC
    451  CTGCAATACA AAATCACCAA ACAGGGCGAA GGCAAACAGC CGACCAAAGA
    501  CGACATCGTT ACCGTGGAAT ACGAAGGCCG CCTGATTGAC GGTACGGTAT
    551  TCGACAGCAG CAAAGCCAAC GGCGGCCCGG TCACCTTCCC TTTGAGCCAA
    601  GTGATTCCGG GTTGGACCGA AGGCGTACAG CTTCTGAAAG AAGGCGGCGA
    651  AGCCACGTTC TACATCCCGT CCAACCTTGC CTACCGCGAA CAGGGTGCGG
    701  GCGACAAAAT CGGTCCGAAC GCCACTTTGG TATTTGATGT GAAACTGGTC
    751  AAAATCGGCG CACCCGAAAA CGCGCCCGCC AAGCAGCCGG CTCAAGTCGA
    801  CATCAAAAAA GTAAATTAA
```

This corresponds to the amino acid sequence <SEQ ID 987; ORF 576-1>:

```
m576-1.pep
      1  MNTIFKISAL TLSAALALSA CGKKEAAPAS ASEPAAASSA QGDTSSIGST
     51  MQQASYAMGV DIGRSLKQMK EQGAEIDLKV FTEAMQAVYD GKEIKMTEEQ
    101  AQEVMMKFLQ EQQAKAVEKH KADAKANKEK GEAFLKENAA KDGVKTTASG
    151  LQYKITKQGE GKQPTKDDIV TVEYEGRLID GTVFDSSKAN GGPVTFPLSQ
    201  VIPGWTEGVQ LLKEGGEATF YIPSNLAYRE QGAGDKIGPN ATLVFDVKLV
    251  KIGAPENAPA KQPAQVDIKK VN*
```

The following DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 988>:

```
g576-1.seq
      1  ATGAACACCA TTTTCAAAAT CAGCGCACTG ACCCTTTCCG CCGCTTTGGC
     51  ACTTTCCGCC TGCGGCAAAA AAGAAGCCGC CCCCGCATCT GCATCCGAAC
    101  CTGCCGCCGC TTCTGCCGCG CAGGGCGACA CCTCTTCAAT CGGCAGCACG
    151  ATGCAGCAGG CAAGCTATGC AATGGGCGTG GACATCGGAC GCTCCCTGAA
```

```
201  ACAAATGAAG GAACAGGGCG CGGAAATCGA TTTGAAAGTC TTTACCGATG
251  CCATGCAGGC AGTGTATGAC GGCAAAGAAA TCAAAATGAC CGAAGAGCAG
301  GCCCAGGAAG TGATGATGAA ATTCCTGCAG GAGCAGCAGG CTAAAGCCGT
351  AGAAAAACAC AAGGCGGATG CGAAGGCCAA CAAAGAAAAA GGCGAAGCCT
401  TCCTGAAGGA AAATGCCGCC AAAGACGGCG TGAAGACCAC TGCTTCCGGT
451  CTGCAGTACA AAATCACCAA ACAGGGTGAA GGCAAACAGC CGACAAAAGA
501  CGACATCGTT ACCGTGGAAT ACGAAGGCCG CCTGATTGAC GGTACCGTAT
551  TCGACAGCAG CAAAGCCAAC GGCGGCCCGG CCACCTTCCC TTTGAGCCAA
601  GTGATTCCGG GTTGGACCGA AGGCGTACGG CTTCTGAAAG AAGGCGGCGA
651  AGCCACGTTC TACATCCCGT CCAACCTTGC CTACCGCGAA CAGGGTGCGG
701  GCGAAAAAAT CGGTCCGAAC GCCACTTTGG TATTTGACGT GAAACTGGTC
751  AAAATCGGCG CACCCGAAAA CGCGCCCGCC AAGCAGCCGG ATCAAGTCGA
801  CATCAAAAAA GTAAATTAA
```

This corresponds to the amino acid sequence <SEQ ID 989; ORF 576-1.ng>:

```
g576-1.pep
  1  MNTIFKISAL TLSAALALSA CGKKEAAPAS ASEPAAASAA QGDTSSIGST
 51  MQQASYAMGV DIGRSLKQMK EQGAEIDLKV FTDAMQAVYD GKEIKMTEEQ
101  AQEVMMKFLQ EQQAKAVEKH KADAKANKEK GEAFLKENAA KDGVKTTASG
151  LQYKITKQGE GKQPTKDDIV TVEYEGRLID GTVFDSSKAN GGPATFPLSQ
201  VIPGWTEGVR LLKEGGEATF YIPSNLAYRE QGAGEKIGPN ATLVFDVKLV
251  KIGAPENAPA KQPDQVDIKK VN*
```

g576-1/m576-1 ORFs 576-1 and 576-1.ng showed a 97.8% identity in 272 aa overlap

```
                      10        20        30        40        50        60
g576-1.pep    MNTIFKISALTLSAALALSACGKKEAAPASASEPAAASAAQGDTSSIGSTMQQASYAMGV
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
m576-1        MNTIFKISALTLSAALALSACGKKEAAPASASEPAAASSAQGDTSSIGSTMQQASYAMGV
                      10        20        30        40        50        60

                      70        80        90       100       110       120
g576-1.pep    DIGRSLKQMKEQGAEIDLKVFTDAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKH
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
m576-1        DIGRSLKQMKEQGAEIDLKVFTEAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKH
                      70        80        90       100       110       120

                     130       140       150       160       170       180
g576-1.pep    KADAKANKEKGEAFLKENAAKDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLID
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m576-1        KADAKANKEKGEAFLKENAAKDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLID
                     130       140       150       160       170       180

                     190       200       210       220       230       240
g576-1.pep    GTVFDSSKANGGPATFPLSQVIPGWTEGVRLLKEGGEATFYIPSNLAYREQGAGEKIGPN
              ||||||||||||:||||||||||||||||:||||||||||||||||||||||||:|||||
m576-1        GTVFDSSKANGGPVTFPLSQVIPGWTEGVQLLKEGGEATFYIPSNLAYREQGAGDKIGPN
                     190       200       210       220       230       240

                     250       260       270
g576-1.pep    ATLVFDVKLVKIGAPENAPAKQPDQVDIKKVNX
              ||||||||||||||||||||||| |||||||||
m576-1        ATLVFDVKLVKIGAPENAPAKQPAQVDIKKVNX
                     250       260       270
```

The following DNA sequence was identified in *N. meningitidis* <SEQ ID 990>:

```
a576-1.seq
  1  ATGAACACCA TTTTCAAAAT CAGCGCACTG ACCCTTTCCG CCGCTTTGGC
 51  ACTTTCCGCC TGCGGCAAAA AAGAAGCCGC CCCCGCATCT GCATCCGAAC
101  CTGCCGCCGC TTCTTCCGCG CAGGGCGACA CCTCTTCGAT CGGCAGCACG
```

```
151  ATGCAGCAGG CAAGCTATGC GATGGGCGTG GACATCGGAC GCTCCCTGAA
201  GCAAATGAAG GAACAGGGCG CGGAAATCGA TTTGAAAGTC TTTACCGAAG
251  CCATGCAGGC AGTGTATGAC GGCAAAGAAA TCAAAATGAC CGAAGAGCAG
301  GCTCAGGAAG TCATGATGAA ATTCCTTCAG GAACAACAGG CTAAAGCCGT
351  AGAAAAACAC AAGGCGGACG CGAAGGCCAA TAAAGAAAAA GGCGAAGCCT
401  TTCTGAAAGA AAATGCCGCC AAAGACGGCG TGAAGACCAC TGCTTCCGGC
451  CTGCAATACA AAATCACCAA ACAGGGCGAA GGCAAACAGC CGACCAAAGA
501  CGACATCGTT ACCGTGGAAT ACGAAGGCCG CCTGATTGAC GGTACGGTAT
551  TCGACAGCAG CAAAGCCAAC GGCGGCCCGG TCACCTTCCC TTTGAGCCAA
601  GTGATTCTGG GTTGGACCGA AGGCGTACAG CTTCTGAAAG AAGGCGGCGA
651  AGCCACGTTC TACATCCCGT CCAACCTTGC CTACCGCGAA CAGGGTGCGG
701  GCGACAAAAT CGGCCCGAAC GCCACTTTGG TATTTGATGT GAAACTGGTC
751  AAAATCGGCG CACCCGAAAA CGCGCCCGCC AAGCAGCCGG CTCAAGTCGA
801  CATCAAAAAA GTAAATTAA
```

This corresponds to the amino acid sequence <SEQ ID 991; ORF 576-1.a>:

```
a576-1.pep
  1  MNTIFKISAL TLSAALALSA CGKKEAAPAS ASEPAAASSA QGDTSSIGST
 51  MQQASYAMGV DIGRSLKQMK EQGAEIDLKV FTEAMQAVYD GKEIKMTEEQ
101  AQEVMMKFLQ EQQAKAVEKH KADAKANKEK GEAFLKENAA KDGVKTTASG
151  LQYKITKQGE GKQPTKDDIV TVEYEGRLID GTVFDSSKAN GGPVTFPLSQ
201  VILGWTEGVQ LLKEGGEATF YIPSNLAYRE QGAGDKIGPN ATLVFDVKLV
251  KIGAPENAPA KQPAQVDIKK VN*
```

a576-1/m576-1 ORFs 576-1 and 576-1.a 99.6% identity in 272 aa overlap

```
                    10        20        30        40        50        60
a576-1.pep  MNTIFKISALTLSAALALSACGKKEAAPASASEPAAASSAQGDTSSIGSTMQQASYAMGV
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m576-1      MNTIFKISALTLSAALALSACGKKEAAPASASEPAAASSAQGDTSSIGSTMQQASYAMGV
                    10        20        30        40        50        60

                    70        80        90       100       110       120
a576-1.pep  DIGRSLKQMKEQGAEIDLKVFTEAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKH
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m576-1      DIGRSLKQMKEQGAEIDLKVFTEAMQAVYDGKEIKMTEEQAQEVMMKFLQEQQAKAVEKH
                    70        80        90       100       110       120

                   130       140       150       160       170       180
a576-1.pep  KADAKANKEKGEAFLKENAAKDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLID
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m576-1      KADAKANKEKGEAFLKENAAKDGVKTTASGLQYKITKQGEGKQPTKDDIVTVEYEGRLID
                   130       140       150       160       170       180

                   190       200       210       220       230       240
a576-1.pep  GTVFDSSKANGGPVTFPLSQVILGWTEGVQLLKEGGEATFYIPSNLAYREQGAGDKIGPN
            |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
m576-1      GTVFDSSKANGGPVTFPLSQVIPGWTEGVQLLKEGGEATFYIPSNLAYREQGAGDKIGPN
                   190       200       210       220       230       240

                   250       260       270
a576-1.pep  ATLVFDVKLVKIGAPENAPAKQPAQVDIKKVNX
            ||||||||||||||||||||||||||||||||
m576-1      ATLVFDVKLVKIGAPENAPAKQPAQVDIKKVNX
                   250       260       270
```

919 and 919-2 gnm43.seq

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 992>:

m919.seq

```
   1  ATGAAAAAAT ACCTATTCCG CGCCGCCCTG TACGGCATCG CCGCCGCCAT
  51  CCTCGCCGCC TGCCAAAGCA AGAGCATCCA AACCTTTCCG CAACCCGACA
 101  CATCCGTCAT CAACGGCCCG GACCGGCCGG TCGGCATCCC CGACCCCGCC
 151  GGAACGACGG TCGGCGGCGG CGGGGCCGTC TATACCGTTG TACCGCACCT
 201  GTCCCTGCCC CACTGGGCGG CGCAGGATTT CGCCAAAAGC CTGCAATCCT
 251  TCCGCCTCGG CTGCGCCAAT TTGAAAAACC GCCAAGGCTG GCAGGATGTG
 301  TGCGCCCAAG CCTTTCAAAC CCCCGTCCAT TCCTTTCAGG CAAAACAGTT
 351  TTTTGAACGC TATTTCACGC CGTGGCAGGT TGCAGGCAAC GGAAGCCTTG
 401  CCGGTACGGT TACCGGCTAT TACGAACCGG TGCTGAAGGG CGACGACAGG
 451  CGGACGGCAC AAGCCCGCTT CCCGATTTAC GGTATTCCCG ACGATTTTAT
 501  CTCCGTCCCC CTGCCTGCCG GTTTGCGGAG CGGAAAAGCC CTTGTCCGCA
 551  TCAGGCAGAC GGGAAAAAAC AGCGGCACAA TCGACAATAC CGGCGGCACA
 601  CATACCGCCG ACCTCTCCcG ATTCCCCATC ACCGCGCGCA CAACAGCAAT
 651  CAAAGGCAGG TTTGAAGGAA GCCGCTTCCT CCCCTACCAC ACGCGCAACC
 701  AAATCAACGG CGGCGCGCTT GACGGCAAAG CCCCGATACT CGGTTACGCC
 751  GAAGACCCTG TCGAACTTTT TTTTATGCAC ATCCAAGGCT CGGGCCGTCT
 801  GAAAACCCCG TCCGGCAAAT ACATCCGCAT CGGCTATGCC GACAAAAACG
 851  AACATCCyTA CGTTTCCATC GGACGCTATA TGGCGGATAA GGGCTACCTC
 901  AAACTCGGAC AAACCTCCAT GCAGGGCATT AAGTCTTATA TGCGGCAAAA
 951  TCCGCAACGC CTCGCCGAAG TTTTGGGTCA AAACCCCAGC TATATCTTTT
1001  TCCGCGAGCT TGCCGGAAGC AGCAATGACG GCCCTGTCGG CGCACTGGGC
1051  ACGCCGCTGA TGGGGGAATA TGCCGGCGCA GTCGACCGGC ACTACATTAC
1101  CTTGGGTGCG CCCTTATTTG TCGCCACCGC CCATCCGGTT ACCCGCAAAG
1151  CCCTCAACCG CCTGATTATG GCGCAGGATA CCGGCAGCGC GATTAAAGGC
1201  GCGGTGCGCG TGGATTATTT TTGGGGATAC GGCGACGAAG CCGGCGAACT
1251  TGCCGGCAAA CAGAAAACCA CGGGATATGT CTGGCAGCTC CTACCCAACG
1301  GTATGAAGCC CGAATACCGc CCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 993; ORF 919>:

m919.pep

```
   1  MKKYLFRAAL YGIAAAILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
  51  GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
 101  CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
 151  RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
 201  HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
 251  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
 301  KLGQTSMQGI KSYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
 351  TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
 401  AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 994>:

m919-2.seq

```
   1  ATGAAAAAAT ACCTATTCCG CGCCGCCCTG TACGGCATCG CCGCCGCCAT
  51  CCTCGCCGCC TGCCAAAGCA AGAGCATCCA AACCTTTCCG CAACCCGACA
 101  CATCCGTCAT CAACGGCCCG GACCGGCCGG TCGGCATCCC CGACCCCGCC
 151  GGAACGACGG TCGGCGGCGG CGGGGCCGTC TATACCGTTG TACCGCACCT
 201  GTCCCTGCCC CACTGGGCGG CGCAGGATTT CGCCAAAAGC CTGCAATCCT
 251  TCCGCCTCGG CTGCGCCAAT TTGAAAAACC GCCAAGGCTG GCAGGATGTG
 301  TGCGCCCAAG CCTTTCAAAC CCCCGTCCAT TCCTTTCAGG CAAAACAGTT
 351  TTTTGAACGC TATTTCACGC CGTGGCAGGT TGCAGGCAAC GGAAGCCTTG
 401  CCGGTACGGT TACCGGCTAT TACGAACCGG TGCTGAAGGG CGACGACAGG
 451  CGGACGGCAC AAGCCCGCTT CCCGATTTAC GGTATTCCCG ACGATTTTAT
 501  CTCCGTCCCC CTGCCTGCCG GTTTGCGGAG CGGAAAAGCC CTTGTCCGCA
 551  TCAGGCAGAC GGGAAAAAAC AGCGGCACAA TCGACAATAC CGGCGGCACA
```

```
 601  CATACCGCCG ACCTCTCCCG ATTCCCCATC ACCGCGCGCA CAACAGCAAT
 651  CAAAGGCAGG TTTGAAGGAA GCCGCTTCCT CCCCTACCAC ACGCGCAACC
 701  AAATCAACGG CGGCGCGCTT GACGGCAAAG CCCCGATACT CGGTTACGCC
 751  GAAGACCCTG TCGAACTTTT TTTTATGCAC ATCCAAGGCT CGGGCCGTCT
 801  GAAAACCCCG TCCGGCAAAT ACATCCGCAT CGGCTATGCC GACAAAAACG
 851  AACATCCCTA CGTTTCCATC GGACGCTATA TGGCGGATAA GGGCTACCTC
 901  AAACTCGGAC AAACCTCCAT GCAGGGCATT AAGTCTTATA TGCGGCAAAA
 951  TCCGCAACGC CTCGCCGAAG TTTTGGGTCA AAACCCCAGC TATATCTTTT
1001  TCCGCGAGCT TGCCGGAAGC AGCAATGACG GCCCTGTCGG CGCACTGGGC
1051  ACGCCGCTGA TGGGGGAATA TGCCGGCGCA GTCGACCGGC ACTACATTAC
1101  CTTGGGTGCG CCCTTATTTG TCGCCACCGC CCATCCGGTT ACCCGCAAAG
1151  CCCTCAACCG CCTGATTATG GCGCAGGATA CCGGCAGCGC GATTAAAGGC
1201  GCGGTGCGCG TGGATTATTT TTGGGGATAC GGCGACGAAG CCGGCGAACT
1251  TGCCGGCAAA CAGAAAACCA CGGGATATGT CTGGCAGCTC CTACCCAACG
1301  GTATGAAGCC CGAATACCGC CCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 995; ORF 919-2>:

m919-2.pep

```
   1  MKKYLFRAAL YGIAAAILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
  51  GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
 101  CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
 151  RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
 201  HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
 251  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
 301  KLGQTSMQGI KSYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
 351  TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
 401  AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

The following partial DNA sequence was identified in *N.gonorrhoeae* <SEQ ID 996>:

g919.seq

```
   1  ATGAAAAAAC ACCTGCTCCG CTCCGCCCTG TACGGcatCG CCGCCgccAT
  51  CctcgCCGCC TGCCAAAgca gGAGCATCCA AACCTTTCCG CAACCCGACA
 101  CATCCGTCAT CAACGGCCCG GACCGGCCGG CCGGCATCCC CGACCCCGCC
 151  GGAACGACGG TTGCCGGCGG CGGGGCCGTC TATACCGTTG TGCCGCACCT
 201  GTCCATGCCC CACTGGGCGG CGCaggATTT TGCCAAAAGC CTGCAATCCT
 251  TCCGCCTCGG CTGCGCCAAT TTGAAAAACC GCCAAGGCTG GCAGGATGTG
 301  TGCGCCCAAG CCTTTCAAAC CCCCGTGCAT TCCTTTCAGG CAAAGcGgTT
 351  TTTTGAACGC TATTTCACGC cgtGGCaggt tgcaggcaAC GGAAGcCTTG
 401  Caggtacggt TACCGGCTAT TACGAACCGG TGCTGAAGGG CGACGGCAGG
 451  CGGACGGAAC GGGCCCGCTT CCCGATTTAC GGTATTCCCG ACGATTTTAT
 501  CTCCGTCCCG CTGCCTGCCG GTTTGCGGGG CGGAAAAAAC CTTGTCCGCA
 551  TCAGGCAGAc ggGGAAAAAC AGCGGCACGA TCGACAATGC CGGCGGCACG
 601  CATACCGCCG ACCTCTCCCG ATTCCCCATC ACCGCGCGCA CAACGGcaat
 651  caaaGGCAGG TTTGAaggAA GCCGCTTCCT CCCTTACCAC ACGCGCAACC
 701  AAAtcaacGG CGGCgcgcTT GACGGCAAag cccCCATCCT CggttacgcC
 751  GAagacccCG tcgaacttTT TTTCATGCAC AtccaaggCT CGGGCCGCCT
 801  GAAAACCCcg tccggcaaat acatCCGCAt cggaTacgcc gacAAAAACG
 851  AACAtccgTa tgtttccatc ggACGctaTA TGGCGGACAA AGGCTACCTC
 901  AAGctcgggc agACCTCGAT GCAGGgcatc aaagcCTATA TGCGGCAAAA
 951  TCCGCAACGC CTCGCCGAAG TTTTGGGTCA AAACCCCAGC TATATCTTTT
1001  TCCGCGAGCT TGCCGGAAGC GGCAATGAGG GCCCCGTCGG CGCACTGGGC
1051  ACGCCACTGA TGGGGGAATA CGCCGGCGCA ATCGACCGGC ACTACATTAC
1101  CTTGGGCGCG CCCTTATTTG TCGCCACCGC CCATCCGGTT ACCCGCAAAG
1151  CCCTCAACCG CCTGATTATG GCGCAGGATA CAGGCAGCGC GATCAAAGGC
1201  GCGGTGCGCG TGGATTATTT TTGGGGTTAC GGCGACGAAG CCGGCGAACT
1251  TGCCGGCAAA CAGAAAACCA CGGGATACGT CTGGCAGCTC CTGCCCAACG
1301  GCATGAAGCC CGAATACCGC CCGTGA
```

This corresponds to the amino acid sequence <SEQ ID 997; ORF 919.ng>:

```
g919.pep
      1  MKKHLLRSAL YGIAAAILAA CQSRSIQTFP QPDTSVINGP DRPAGIPDPA
     51  GTTVAGGGAV YTVVPHLSMP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
    101  CAQAFQTPVH SFQAKRFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDGR
    151  RTERARFPIY GIPDDFISVP LPAGLRGGKN LVRIRQTGKN SGTIDNAGGT
    201  HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
    251  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
    301  KLGQTSMQGI KAYMRQNPQR LAEVLGQNPS YIFFRELAGS GNEGPVGALG
    351  TPLMGEYAGA IDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
    401  AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

ORF 919 shows 95.9 % identity over a 441 aa overlap with a predicted ORF (ORF 919.ng) from *N. gonorrhoeae:*

```
m919/g919

                     10        20        30        40        50        60
m919.pep     MKKYLFRAALYGIAAAILAACQSKSIQTFPQPDTSVINGPDRPVGIPDPAGTTVGGGGAV
             |||:|:|:|||||||||||||||:||||||||||||||||||:|||||||||||:|||||
g919         MKKHLLRSALYGIAAAILAACQSRSIQTFPQPDTSVINGPDRPAGIPDPAGTTVAGGGAV
                     10        20        30        40        50        60

                     70        80        90       100       110       120
m919.pep     YTVVPHLSLPHWAAQDFAKSLQSFRLGCANLKNRQGWQDVCAQAFQTPVHSFQAKQFFER
             ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||:|||
g919         YTVVPHLSMPHWAAQDFAKSLQSFRLGCANLKNRQGWQDVCAQAFQTPVHSFQAKRFFER
                     70        80        90       100       110       120

                    130       140       150       160       170       180
m919.pep     YFTPWQVAGNGSLAGTVTGYYEPVLKGDDRRTAQARFPIYGIPDDFISVPLPAGLRSGKA
             |||||||||||||||||||||||||||| ||| :||||||||||||||||||||||:|| 
g919         YFTPWQVAGNGSLAGTVTGYYEPVLKGDGRRTERARFPIYGIPDDFISVPLPAGLRGGKN
                    130       140       150       160       170       180

                    190       200       210       220       230       240
m919.pep     LVRIRQTGKNSGTIDNTGGTHTADLSRFPITARTTAIKGRFEGSRFLPYHTRNQINGGAL
             ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
g919         LVRIRQTGKNSGTIDNAGGTHTADLSRFPITARTTAIKGRFEGSRFLPYHTRNQINGGAL
                    190       200       210       220       230       240

                    250       260       270       280       290       300
m919.pep     DGKAPILGYAEDPVELFFMHIQGSGRLKTPSGKYIRIGYADKNEHPYVSIGRYMADKGYL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
g919         DGKAPILGYAEDPVELFFMHIQGSGRLKTPSGKYIRIGYADKNEHPYVSIGRYMADKGYL
                    250       260       270       280       290       300

                    310       320       330       340       350       360
m919.pep     KLGQTSMQGIKSYMRQNPQRLAEVLGQNPSYIFFRELAGSSNDGPVGALGTPLMGEYAGA
             |||||||||||:|||||||||||||||||||||||||||||:|:||||||||||||||||
g919         KLGQTSMQGIKAYMRQNPQRLAEVLGQNPSYIFFRELAGSGNEGPVGALGTPLMGEYAGA
                    310       320       330       340       350       360

                    370       380       390       400       410       420
m919.pep     VDRHYITLGAPLFVATAHPVTRKALNRLIMAQDTGSAIKGAVRVDYFWGYGDEAGELAGK
             :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
g919         IDRHYITLGAPLFVATAHPVTRKALNRLIMAQDTGSAIKGAVRVDYFWGYGDEAGELAGK
                    370       380       390       400       410       420
```

```
                    430       440
m919.pep     QKTTGYVWQLLPNGMKPEYRPX
             ||||||||||||||||||||||
g919         QKTTGYVWQLLPNGMKPEYRPX
                    430       440
```

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 998>:

```
a919.seq
   1  ATGAAAAAAT ACCTATTCCG CGCCGCCCTG TGCGGCATCG CCGCCGCCAT
  51  CCTCGCCGCC TGCCAAAGCA AGAGCATCCA AACCTTTCCG CAACCCGACA
 101  CATCCGTCAT CAACGGCCCG GACCGGCCGG TCGGCATCCC CGACCCCGCC
 151  GGAACGACGG TCGGCGGCGG CGGGGCCGTT TATACCGTTG TGCCGCACCT
 201  GTCCCTGCCC CACTGGGCGG CGCAGGATTT CGCCAAAAGC CTGCAATCCT
 251  TCCGCCTCGG CTGCGCCAAT TTGAAAAACC GCCAAGGCTG GCAGGATGTG
 301  TGCGCCCAAG CCTTTCAAAC CCCCGTCCAT TCCGTTCAGG CAAAACAGTT
 351  TTTTGAACGC TATTTCACGC CGTGGCAGGT TGCAGGCAAC GGAAGCCTTG
 401  CCGGTACGGT TACCGGCTAT TACGAGCCGG TGCTGAAGGG CGACGACAGG
 451  CGGACGGCAC AAGCCCGCTT CCCGATTTAC GGTATTCCCG ACGATTTTAT
 501  CTCCGTCCCC CTGCCTGCCG GTTTGCGGAG CGGAAAAGCC CTTGTCCGCA
 551  TCAGGCAGAC GGGAAAAAAC AGCGGCACAA TCGACAATAC CGGCGGCACA
 601  CATACCGCCG ACCTCTCCCA ATTCCCCATC ACTGCGCGCA CAACGGCAAT
 651  CAAAGGCAGG TTTGAAGGAA GCCGCTTCCT CCCCTACCAC ACGCGCAACC
 701  AAATCAACGG CGGCGCGCTT GACGGCAAAG CCCCGATACT CGGTTACGCC
 751  GAAGACCCCG TCGAACTTTT TTTTATGCAC ATCCAAGGCT CGGGCCGTCT
 801  GAAAACCCCG TCCGGCAAAT ACATCCGCAT CGGCTATGCC GACAAAAACG
 851  AACATCCCTA CGTTTCCATC GGACGCTATA TGGCGGACAA AGGCTACCTC
 901  AAGCTCGGGC AGACCTCGAT GCAGGGCATC AAAGCCTATA TGCAGCAAAA
 951  CCCGCAACGC CTCGCCGAAG TTTTGGGGCA AAACCCCAGC TATATCTTTT
1001  TCCGAGAGCT TACCGGAAGC AGCAATGACG GCCCTGTCGG CGCACTGGGC
1051  ACGCCGCTGA TGGGCGAGTA CGCCGGCGCA GTCGACCGGC ACTACATTAC
1101  CTTGGGCGCG CCCTTATTTG TCGCCACCGC CCATCCGGTT ACCCGCAAAG
1151  CCCTCAACCG CCTGATTATG GCGCAGGATA CCGGCAGCGC GATTAAAGGC
1201  GCGGTGCGCG TGGATTATTT TTGGGGATAC GGCGACGAAG CCGGCGAACT
1251  TGCCGGCAAA CAGAAAACCA CGGGATATGT CTGGCAGCTT CTGCCCAACG
1301  GTATGAAGCC CGAATACCGC CCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 999; ORF 919.a>:

```
a919.pep
   1  MKKYLFRAAL CGIAAAILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
  51  GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
 101  CAQAFQTPVH SVQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
 151  RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
 201  HTADLSQFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
 251  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
 301  KLGQTSMQGI KAYMQQNPQR LAEVLGQNPS YIFFRELTGS SNDGPVGALG
 351  TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
 401  AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

m919/a919 ORFs 919 and 919.a showed a 98.6% identity in 441 aa overlap

```
                    10        20        30        40        50        60
m919.pep     MKKYLFRAALYGIAAAILAACQSKSIQTFPQPDTSVINGPDRPVGIPDPAGTTVGGGGAV
             |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
a919         MKKYLFRAALCGIAAAILAACQSKSIQTFPQPDTSVINGPDRPVGIPDPAGTTVGGGGAV
                    10        20        30        40        50        60

                    70        80        90       100       110       120
m919.pep     YTVVPHLSLPHWAAQDFAKSLQSFRLGCANLKNRQGWQDVCAQAFQTPVHSFQAKQFFER
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
a919         YTVVPHLSLPHWAAQDFAKSLQSFRLGCANLKNRQGWQDVCAQAFQTPVHSVQAKQFFER
```

```
                    70        80        90       100       110       120

                   130       140       150       160       170       180
m919.pep  YFTPWQVAGNGSLAGTVTGYYEPVLKGDDRRTAQARFPIYGIPDDFISVPLPAGLRSGKA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a919      YFTPWQVAGNGSLAGTVTGYYEPVLKGDDRRTAQARFPIYGIPDDFISVPLPAGLRSGKA
                   130       140       150       160       170       180

                   190       200       210       220       230       240
m919.pep  LVRIRQTGKNSGTIDNTGGTHTADLSRFPITARTTAIKGRFEGSRFLPYHTRNQINGGAL
          |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
a919      LVRIRQTGKNSGTIDNTGGTHTADLSQFPITARTTAIKGRFEGSRFLPYHTRNQINGGAL
                   190       200       210       220       230       240

                   250       260       270       280       290       300
m919.pep  DGKAPILGYAEDPVELFFMHIQGSGRLKTPSGKYIRIGYADKNEHPYVSIGRYMADKGYL
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a919      DGKAPILGYAEDPVELFFMHIQGSGRLKTPSGKYIRIGYADKNEHPYVSIGRYMADKGYL
                   250       260       270       280       290       300

                   310       320       330       340       350       360
m919.pep  KLGQTSMQGIKSYMRQNPQRLAEVLGQNPSYIFFRELAGSSNDGPVGALGTPLMGEYAGA
          |||||||||||:||:|||||||||||||||||||||||:|||||||||||||||||||||
a919      KLGQTSMQGIKAYMQQNPQRLAEVLGQNPSYIFFRELTGSSNDGPVGALGTPLMGEYAGA
                   310       320       330       340       350       360

                   370       380       390       400       410       420
m919.pep  VDRHYITLGAPLFVATAHPVTRKALNRLIMAQDTGSAIKGAVRVDYFWGYGDEAGELAGK
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a919      VDRHYITLGAPLFVATAHPVTRKALNRLIMAQDTGSAIKGAVRVDYFWGYGDEAGELAGK
                   370       380       390       400       410       420

                   430       440
m919.pep  QKTTGYVWQLLPNGMKPEYRPX
          ||||||||||||||||||||||
a919      QKTTGYVWQLLPNGMKPEYRPX
                   430       440
```

121 and 121-1

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1000>:

m121.seq

```
  1  ATGGAAACAC AGCTTTACAT CGGCATCATG TCGGGAACCA GCATGGACGG
 51  GGCGGATGCC GTACTGATAC GGATGGACGG CGGCAAATGG CTGGGCGCGG
101  AAGGGCACGC CTTTACCCCC TACCCCGGCA GGTTACGCCG CCAATTGCTG
151  GATTTGCAGG ACACAGGCGC AGACGAACTG CACCGCAGCA GGATTTTGTC
201  GCAAGAACTC AGCCGCCTAT ATGCGCAAAC CGCCGCCGAA CTGCTGTGCA
251  GTCAAAACCT CGCACCGTCC GACATTACCG CCCTCGGCTG CCACGGGCAA
301  ACCGTCCGAC ACGCGCCGGA ACACGGTTAC AGCATACAGC TTGCCGATTT
351  GCCGCTGCTG GCGxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
401  xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
451  xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
501  xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
551  xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
601  xxxxxxCAGC TTCCTTACGA CAAAAACGGT GCAAAGTCGG CACAAGGCAA
651  CATATTGCCG CAACTGCTCG ACAGGCTGCT CGCCCACCCG TATTTCGCAC
701  AACGCCACCC TAAAAGCACG GGGCGCGAAC TGTTTGCCAT AAATTGGCTC
751  GAAACCTACC TTGACGGCGG CGAAAACCGA TACGACGTAT TGCGGACGCT
801  TTCCCGTTTT ACCGCGCAAA CCGTTTGCGA CGCCGTCTCA CACGCAGCGG
```

```
 851  CAGATGCCCG TCAAATGTAC ATTTGCGACG GCGGCATCCG CAATCCTGTT
 901  TTAATGGCGG ATTTGGCAGA ATGTTTCGGC ACACGCGTTT CCCTGCACAG
 951  CACCGCCGAC CTGAACCTCG ATCCGCAATG GGTGGAAGCC GCCGnATTTG
1001  CGTGGTTGGC GGCGTGTTGG ATTAATCGCA TTCCCGGTAG TCCGCACAAA
1051  GCAACCGGCG CATCCAAACC GTGTATTCTG AnCGCGGGAT ATTATTATTG
1101  A
```

This corresponds to the amino acid sequence <SEQ ID 1001; ORF 121>:

m121.pep

```
   1  METQLYIGIM SGTSMDGADA VLIRMDGGKW LGAEGHAFTP YPGRLRRQLL
  51  DLQDTGADEL HRSRILSQEL SRLYAQTAAE LLCSQNLAPS DITALGCHGQ
 101  TVRHAPEHGY SIQLADLPLL Axxxxxxxxx xxxxxxxxxx xxxxxxxxxx
 151  xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx
 201  xxQLPYDKNG AKSAQGNILP QLLDRLLAHP YFAQRHPKST GRELFAINWL
 251  ETYLDGGENR YDVLRTLSRF TAQTVCDAVS HAAADARQMY ICDGGIRNPV
 301  LMADLAECFG TRVSLHSTAD LNLDPQWVEA AXFAWLAACW INRIPGSPHK
 351  ATGASKPCIL XAGYYY*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1002>:

g121.seq

```
   1  ATGGAAACAC AGCTTTACAT CGGCATTATG TCGGGAACCA GTATGGACGG
  51  GGCGGATGCC GTGCTGGTAC GGATGGACGG CGGCAAATGG CTGGGCGCGG
 101  AAGGGCACGC CTTTACCCCC TACCCTGACC GGTTGCGCCG CAAATTGCTG
 151  GATTTGCAGG ACACAGGCAC AGACGAACTG CACCGCAGCA GGATGTTGTC
 201  GCAAGAACTC AGCCGCCTGT ACGCGCAAAC CGCCGCCGAA CTGCTGTGCA
 251  GTCAAAACCT CGCTCCGTGC GACATTACCG CCCTCGGCTG CCACGGGCAA
 301  ACCGTCCGAC ACGCGCCGGA ACACGGTtac AGCATACAGC TTGCCGATTT
 351  GCCGCTGCTG GCGGAACTGa cgcggatttT TACCGTCggc gacttcCGCA
 401  GCCGCGACCT TGCTGCCGGC GGacaAGGTG CGCCGCTCGT CCCCGCCTTT
 451  CACGAAGCCC TGTTCCGCGA TGACAGGGAA ACACGCGTGG TACTGAACAT
 501  CGGCGGGATT GCCAACATCA GCGTACTCCC CCCCGGCGCA CCCGCCTTCG
 551  GCTTCGACAC AGGGCCGGGC AATATGCTGA TGGAcgcgtg gacgcaggca
 601  cacTGGcagc TGCCTTACGA CAAAAacggt gcAAAGgcgg cacAAGGCAA
 651  catatTGCcg cAACTGCTCG gcaggctGCT CGCCcacccG TATTTCTCAC
 701  AACCCcaccc aaAAAGCACG GGcGCGaac TgtttgcccT AAattggctc
 751  gaaacctAcc ttgacggcgg cgaaaaccga tacgacgtat tgcggacgct
 801  ttcccgattc accgcgcaaA ccgTttggga cgccgtctca CACGCAGCGG
 851  CAGATGCCCG TCAAATGTAC ATTTGCGGCG GCGGCATCCG CAATCCTGTT
 901  TTAATGGCGG ATTTGGCAGA ATGTTTCGGC ACACGCGTTT CCCTGCACAG
 951  CACCGCCGAA CTGAACCTCG ATCCTCAATG GGTGGAGGCG gccgCATTtg
1001  cgtggttggC GGCGTGTTGG ATTAACCGCA TTCCCGGTAG TCCGCACAAA
1051  GCGACCGGCG CATCCAAACC GTGTATTCTG GGCGCGGGAT ATTATTATTG
1101  A
```

This corresponds to the amino acid sequence <SEQ ID 1003; ORF 121.ng>:

g121.pep

```
   1  METQLYIGIM SGTSMDGADA VLVRMDGGKW LGAEGHAFTP YPDRLRRKLL
  51  DLQDTGTDEL HRSRMLSQEL SRLYAQTAAE LLCSQNLAPC DITALGCHGQ
 101  TVRHAPEHGY SIQLADLPLL AELTRIFTVG DFRSRDLAAG GQGAPLVPAF
 151  HEALFRDDRE TRVVLNIGGI ANISVLPPGA PAFGFDTGPG NMLMDAWTQA
 201  HWQLPYDKNG AKAAQGNILP QLLGRLLAHP YFSQPHPKST GRELFALNWL
 251  ETYLDGGENR YDVLRTLSRF TAQTVWDAVS HAAADARQMY ICGGGIRNPV
 301  LMADLAECFG TRVSLHSTAE LNLDPQWVEA AAFAWLAACW INRIPGSPHK
 351  ATGASKPCIL GAGYYY*
```

ORF 121 shows 73.5% identity over a 366 aa overlap with a predicted ORF (ORF121.ng) from *N. gonorrhoeae:*

m121/g121

```
                    10        20        30        40        50        60
m121.pep    METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRQLLDLQDTGADEL
            ||||||||||||||||||||||:||||||||||||||||||| |||||:||||||||:|||
g121        METQLYIGIMSGTSMDGADAVLVRMDGGKWLGAEGHAFTPYPDRLRRKLLDLQDTGTDEL
                    10        20        30        40        50        60
                    70        80        90       100       110       120
m121.pep    HRSRILSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHGYSIQLADLPLL
            ||||:|||||||||||||||||||||||| ||||||||||||||||||||||||||||||
g121        HRSRMLSQELSRLYAQTAAELLCSQNLAPCDITALGCHGQTVRHAPEHGYSIQLADLPLL
                    70        80        90       100       110       120
                   130       140       150       160       170       180
m121.pep    AXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
            | :  :                                  :
g121        AELTRIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDDRETRVVLNIGGIANISVLPPGA
                   130       140       150       160       170       180
                   190       200       210       220       230       240
m121.pep    XXXXXXXXXXXXXXXXXXXXXXQLPYDKNGAKSAQGNILPQLLDRLLAHPYFAQRHPKST
                  :          :    ||||||||||:|||||||||| ||||||||:| |||||
g121        PAFGFDTGPGNMLMDAWTQAHWQLPYDKNGAKAAQGNILPQLLGRLLAHPYFSQPHPKST
                   190       200       210       220       230       240
                   250       260       270       280       290       300
m121.pep    GRELFAINWLETYLDGGENRYDVLRTLSRFTAQTVCDAVSHAAADARQMYICDGGIRNPV
            ||||||:|||||||||||||||||||||||||||| |||||||||||||||| |||||||
g121        GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVWDAVSHAAADARQMYICGGGIRNPV
                   250       260       270       280       290       300
                   310       320       330       340       350       360
m121.pep    LMADLAECFGTRVSLHSTADLNLDPQWVEAAXFAWLAACWINRIPGSPHKATGASKPCIL
            |||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||
g121        LMADLAECFGTRVSLHSTAELNLDPQWVEAAAFAWLAACWINRIPGSPHKATGASKPCIL
                   310       320       330       340       350       360

m121.pep    XAGYYYX
             ||||||
g121        GAGYYYX
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1004>:

```
a121.seq
   1  ATGGAAACAC AGCTTTACAT CGGCATCATG TCGGGAACCA GCATGGACGG
  51  GGCGGATGCC GTACTGATAC GGATGGACGG CGGCAAATGG CTGGGCGCGG
 101  AAGGGCACGC CTTTACCCCC TACCCCGGCA GGTTACGCCG CAAATTGCTG
 151  GATTTGCAGG ACACAGGCGC GGACGAACTG CACCGCAGCA GGATGTTGTC
 201  GCAAGAACTC AGCCGCCTGT ACGCGCAAAC CGCCGCCGAA CTGCTGTGCA
 251  GTCAAAACCT CGCGCCGTCC GACATTACCG CCCTCGGCTG CCACGGGCAA
 301  ACCGTCAGAC ACGCGCCGGA ACACAGTTAC AGCGTACAGC TTGCCGATTT
 351  GCCGCTGCTG GCGGAACGGA CTCAGATTTT TACCGTCGGC GACTTCCGCA
 401  GCCGCGACCT TGCGGCCGGC GGACAAGGCG CGCCGCTCGT CCCCGCCTTT
 451  CACGAAGCCC TGTTCCGCGA CGACAGGGAA ACACGCGCGG TACTGAACAT
 501  CGGCGGGATT GCCAACATCA GCGTACTCCC CCCCGACGCA CCCGCCTTCG
 551  GCTTCGACAC AGGACCGGGC AATATGCTGA TGGACGCGTG GATGCAGGCA
 601  CACTGGCAGC TTCCTTACGA CAAAAACGGT GCAAAGGCGG CACAAGGCAA
 651  CATATTGCCG CAACTGCTCG ACAGGCTGCT CGCCCACCCG TATTTCGCAC
 701  AACCCCACCC TAAAAGCACG GGGCGCGAAC TGTTTGCCCT AAATTGGCTC
 751  GAAACCTACC TTGACGGCGG CGAAAACCGA TACGACGTAT TGCGGACGCT
 801  TTCCCGATTC ACCGCGCAAA CCGTTTTCGA CGCCGTCTCA CACGCAGCGG
 851  CAGATGCCCG TCAAATGTAC ATTTGCGGCG GCGGCATCCG CAATCCTGTT
 901  TTAATGGCGG ATTTGGCAGA ATGTTTCGGC ACACGCGTTT CCCTGCACAG
 951  CACCGCCGAA CTGAACCTCG ATCCGCAATG GGTAGAAGCC GCCGCGTTCG
1001  CATGGATGGC GGCGTGTTGG GTCAACCGCA TTCCCGGTAG TCCGCACAAA
1051  GCAACCGGCG CATCCAAACC GTGTATTCTG GGCGCGGGAT ATTATTATTG
1101  A
```

This corresponds to the amino acid sequence <SEQ ID 1005; ORF 121.a>:

```
a121.pep
      1  METQLYIGIM SGTSMDGADA VLIRMDGGKW LGAEGHAFTP YPGRLRRKLL
     51  DLQDTGADEL HRSRMLSQEL SRLYAQTAAE LLCSQNLAPS DITALGCHGQ
    101  TVRHAPEHSY SVQLADLPLL AERTQIFTVG DFRSRDLAAG GQGAPLVPAF
    151  HEALFRDDRE TRAVLNIGGI ANISVLPPDA PAFGFDTGPG NMLMDAWMQA
    201  HWQLPYDKNG AKAAQGNILP QLLDRLLAHP YFAQPHPKST GRELFALNWL
    251  ETYLDGGENR YDVLRTLSRF TAQTVFDAVS HAAADARQMY ICGGGIRNPV
    301  LMADLAECFG TRVSLHSTAE LNLDPQWVEA AAFAWMAACW VNRIPGSPHK
    351  ATGASKPCIL GAGYYY*
```

m121/a121 **ORFs 121 and 121.a** 74.0% identity in 366 aa overlap

```
                    10        20        30        40        50        60
m121.pep    METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRQLLDLQDTGADEL
            |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
a121        METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRKLLDLQDTGADEL
                    10        20        30        40        50        60

                    70        80        90       100       110       120
m121.pep    HRSRILSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHGYSIQLADLPLL
            ||||:|||||||||||||||||||||||||||||||||||||||||||:||:||||||||
a121        HRSRMLSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHSYSVQLADLPLL
                    70        80        90       100       110       120

                   130       140       150       160       170       180
m121.pep    AXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
            |  :    :                                :
a121        AERTQIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDDRETRAVLNIGGIANISVLPPDA
                   130       140       150       160       170       180

                   190       200       210       220       230       240
m121.pep    XXXXXXXXXXXXXXXXXXXXXXQLPYDKNGAKSAQGNILPQLLDRLLAHPYFAQRHPKST
                 :                ||||||||:||||||||||||||||||||| ||||||
a121        PAFGFDTGPGNMLMDAWMQAHWQLPYDKNGAKAAQGNILPQLLDRLLAHPYFAQPHPKST
                   190       200       210       220       230       240

                   250       260       270       280       290       300
m121.pep    GRELFAINWLETYLDGGENRYDVLRTLSRFTAQTVCDAVSHAAADARQMYICDGGIRNPV
            ||||||:||||||||||||||||||||||| ||||||||||||||| |||||||
a121        GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVFDAVSHAAADARQMYICGGGIRNPV
                   250       260       270       280       290       300

                   310       320       330       340       350       360
m121.pep    LMADLAECFGTRVSLHSTADLNLDPQWVEAAXFAWLAACWINRIPGSPHKATGASKPCIL
            ||||||||||||||||||:|||||||||||| |||:||||:|||||||||||||||||||
a121        LMADLAECFGTRVSLHSTAELNLDPQWVEAAAFAWMAACWVNRIPGSPHKATGASKPCIL
                   310       320       330       340       350       360

m121.pep    XAGYYYX
             ||||||
a121        GAGYYYX
```

Further work revealed the DNA sequence identified in *N. meningitidis* <SEQ ID 1006>:

```
m121-1.seq
      1  ATGGAAACAC AGCTTTACAT CGGCATCATG TCGGGAACCA GCATGGACGG
     51  GGCGGATGCC GTACTGATAC GGATGGACGG CGGCAAATGG CTGGGCGCGG
    101  AAGGGCACGC CTTTACCCCC TACCCCGGCA GGTTACGCCG CCAATTGCTG
    151  GATTTGCAGG ACACAGGCGC AGACGAACTG CACCGCAGCA GGATTTTGTC
    201  GCAAGAACTC AGCCGCCTAT ATGCGCAAAC CGCCGCCGAA CTGCTGTGCA
```

```
 251  GTCAAAACCT CGCACCGTCC GACATTACCG CCCTCGGCTG CCACGGGCAA
 301  ACCGTCCGAC ACGCGCCGGA ACACGGTTAC AGCATACAGC TTGCCGATTT
 351  GCCGCTGCTG GCGGAACGGA CGCGGATTTT TACCGTCGGC GACTTCCGCA
 401  GCCGCGACCT TGCGGCCGGC GGACAAGGCG CGCCACTCGT CCCCGCCTTT
 451  CACGAAGCCC TGTTCCGCGA CAACAGGGAA ACACGCGCGG TACTGAACAT
 501  CGGCGGGATT GCCAACATCA GCGTACTCCC CCCCGACGCA CCCGCCTTCG
 551  GCTTCGACAC AGGGCCGGGC AATATGCTGA TGGACGCGTG GACGCAGGCA
 601  CACTGGCAGC TTCCTTACGA CAAAAACGGT GCAAAGGCGG CACAAGGCAA
 651  CATATTGCCG CAACTGCTCG ACAGGCTGCT CGCCCACCCG TATTTCGCAC
 701  AACCCCACCC TAAAAGCACG GGGCGCGAAC TGTTTGCCCT AAATTGGCTC
 751  GAAACCTACC TTGACGGCGG CGAAAACCGA TACGACGTAT TGCGGACGCT
 801  TTCCCGTTTT ACCGCGCAAA CCGTTTGCGA CGCCGTCTCA CACGCAGCGG
 851  CAGATGCCCG TCAAATGTAC ATTTGCGGCG GCGGCATCCG CAATCCTGTT
 901  TTAATGGCGG ATTTGGCAGA ATGTTTCGGC ACACGCGTTT CCCTGCACAG
 951  CACCGCCGAC CTGAACCTCG ATCCGCAATG GGTGGAAGCC GCCGNATTTG
1001  CGTGGTTGGC GGCGTGTTGG ATTAATCGCA TTCCCGGTAG TCCGCACAAA
1051  GCAACCGGCG CATCCAAACC GTGTATTCTG ANCGCGGGAT ATTATTATTG
1101  A
```

This corresponds to the amino acid sequence <SEQ ID 1007; ORF 121-1>:

```
m121-1.pep
   1  METQLYIGIM SGTSMDGADA VLIRMDGGKW LGAEGHAFTP YPGRLRRQLL
  51  DLQDTGADEL HRSRILSQEL SRLYAQTAAE LLCSQNLAPS DITALGCHGQ
 101  TVRHAPEHGY SIQLADLPLL AERTRIFTVG DFRSRDLAAG GQGAPLVPAF
 151  HEALFRDNRE TRAVLNIGGI ANISVLPPDA PAFGFDTGPG NMLMDAWTQA
 201  HWQLPYDKNG AKAAQGNILP QLLDRLLAHP YFAQPHPKST GRELFALNWL
 251  ETYLDGGENR YDVLRTLSRF TAQTVCDAVS HAAADARQMY ICGGGIRNPV
 301  LMADLAECFG TRVSLHSTAD LNLDPQWVEA AXFAWLAACW INRIPGSPHK
 351  ATGASKPCIL XAGYYY*
```

m121-1/g121 **ORFs 121-1 and 121-1.ng showed a** 95.6% identity in 366 aa overlap

```
                      10        20        30        40        50        60
m121-1.pep  METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRQLLDLQDTGADEL
            ||||||||||||||||||||||:|||||||||||||||||||| ||||:|||||||:|||
g121        METQLYIGIMSGTSMDGADAVLVRMDGGKWLGAEGHAFTPYPDRLRRKLLDLQDTGTDEL
                      10        20        30        40        50        60

                      70        80        90       100       110       120
m121-1.pep  HRSRILSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHGYSIQLADLPLL
            ||||:|||||||||||||||||||||||| ||||||||||||||||||||||||||||||
g121        HRSRMLSQELSRLYAQTAAELLCSQNLAPCDITALGCHGQTVRHAPEHGYSIQLADLPLL
                      70        80        90       100       110       120

                     130       140       150       160       170       180
m121-1.pep  AERTRIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDNRETRAVLNIGGIANISVLPPDA
            || ||||||||||||||||||||||||||||||||||:||||:||||||||||||||| |
g121        AELTRIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDDRETRVVLNIGGIANISVLPPGA
                     130       140       150       160       170       180

                     190       200       210       220       230       240
m121-1.pep  PAFGFDTGPGNMLMDAWTQAHWQLPYDKNGAKAAQGNILPQLLDRLLAHPYFAQPHPKST
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||:|||||||
g121        PAFGFDTGPGNMLMDAWTQAHWQLPYDKNGAKAAQGNILPQLLGRLLAHPYFSQPHPKST
                     190       200       210       220       230       240

                     250       260       270       280       290       300
m121-1.pep  GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVCDAVSHAAADARQMYICGGGIRNPV
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
g121        GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVWDAVSHAAADARQMYICGGGIRNPV
                     250       260       270       280       290       300
```

```
                   310       320       330       340       350       360
m121-1.pep   LMADLAECFGTRVSLHSTADLNLDPQWVEAAXFAWLAACWINRIPGSPHKATGASKPCIL
             |||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||
g121         LMADLAECFGTRVSLHSTAELNLDPQWVEAAAFAWLAACWINRIPGSPHKATGASKPCIL
                   310       320       330       340       350       360

m121-1.pep   XAGYYYX
              ||||||
g121         GAGYYYX
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1008>:

```
a121-1.seq
     1  ATGGAAACAC AGCTTTACAT CGGCATCATG TCGGGAACCA GCATGGACGG
    51  GGCGGATGCC GTACTGATAC GGATGGACGG CGGCAAATGG CTGGGCGCGG
   101  AAGGGCACGC CTTTACCCCC TACCCCGGCA GGTTACGCCG CAAATTGCTG
   151  GATTTGCAGG ACACAGGCGC GGACGAACTG CACCGCAGCA GGATGTTGTC
   201  GCAAGAACTC AGCCGCCTGT ACGCGCAAAC CGCCGCCGAA CTGCTGTGCA
   251  GTCAAAACCT CGCGCCGTCC GACATTACCG CCCTCGGCTG CCACGGGCAA
   301  ACCGTCAGAC ACGCGCCGGA ACACAGTTAC AGCGTACAGC TTGCCGATTT
   351  GCCGCTGCTG GCGGAACGGA CTCAGATTTT TACCGTCGGC GACTTCCGCA
   401  GCCGCGACCT TGCGGCCGGC GGACAAGGCG CGCCGCTCGT CCCCGCCTTT
   451  CACGAAGCCC TGTTCCGCGA CGACAGGGAA ACACGCGCGG TACTGAACAT
   501  CGGCGGGATT GCCAACATCA GCGTACTCCC CCCCGACGCA CCCGCCTTCG
   551  GCTTCGACAC AGGACCGGGC AATATGCTGA TGGACGCGTG GATGCAGGCA
   601  CACTGGCAGC TTCCTTACGA CAAAAACGGT GCAAAGGCGG CACAAGGCAA
   651  CATATTGCCG CAACTGCTCG ACAGGCTGCT CGCCCACCCG TATTTCGCAC
   701  AACCCCACCC TAAAAGCACG GGGCGCGAAC TGTTTGCCCT AAATTGGCTC
   751  GAAACCTACC TTGACGGCGG CGAAAACCGA TACGACGTAT TGCGGACGCT
   801  TTCCCGATTC ACCGCGCAAA CCGTTTTCGA CGCCGTCTCA CACGCAGCGG
   851  CAGATGCCCG TCAAATGTAC ATTTGCGGCG GCGGCATCCG CAATCCTGTT
   901  TTAATGGCGG ATTTGGCAGA ATGTTTCGGC ACACGCGTTT CCCTGCACAG
   951  CACCGCCGAA CTGAACCTCG ATCCGCAATG GGTAGAAGCC GCCGCGTTCG
  1001  CATGGATGGC GGCGTGTTGG GTCAACCGCA TTCCCGGTAG TCCGCACAAA
  1051  GCAACCGGCG CATCCAAACC GTGTATTCTG GGCGCGGGAT ATTATTATTG
  1101  A
```

This corresponds to the amino acid sequence <SEQ ID 1009; ORF 121-1.a>:

```
a121-1.pep
     1  METQLYIGIM SGTSMDGADA VLIRMDGGKW LGAEGHAFTP YPGRLRRKLL
    51  DLQDTGADEL HRSRMLSQEL SRLYAQTAAE LLCSQNLAPS DITALGCHGQ
   101  TVRHAPEHSY SVQLADLPLL AERTQIFTVG DFRSRDLAAG GQGAPLVPAF
   151  HEALFRDDRE TRAVLNIGGI ANISVLPPDA PAFGFDTGPG NMLMDAWMQA
   201  HWQLPYDKNG AKAAQGNILP QLLDRLLAHP YFAQPHPKST GRELFALNWL
   251  ETYLDGGENR YDVLRTLSRF TAQTVFDAVS HAAADARQMY ICGGGIRNPV
   301  LMADLAECFG TRVSLHSTAE LNLDPQWVEA AAFAWMAACW VNRIPGSPHK
   351  ATGASKPCIL GAGYYY*
```

m121-1/a121-1 ORFs 121-1 and 121-1.a showed a 96.4% identity in 366 aa overlap

```
                   10        20        30        40        50        60
m121-1.pep   METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRQLLDLQDTGADEL
             ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
a121-1       METQLYIGIMSGTSMDGADAVLIRMDGGKWLGAEGHAFTPYPGRLRRKLLDLQDTGADEL
                   10        20        30        40        50        60

                   70        80        90       100       110       120
m121-1.pep   HRSRILSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHGYSIQLADLPLL
             ||||:|||||||||||||||||||||||||||||||||||||||||||:||:||||||||
a121-1       HRSRMLSQELSRLYAQTAAELLCSQNLAPSDITALGCHGQTVRHAPEHSYSVQLADLPLL
                   70        80        90       100       110       120
```

```
                 130       140       150       160       170       180
m121-1.pep  AERTRIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDNRETRAVLNIGGIANISVLPPDA
            ||||:||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
a121-1      AERTQIFTVGDFRSRDLAAGGQGAPLVPAFHEALFRDDRETRAVLNIGGIANISVLPPDA
                 130       140       150       160       170       180

                 190       200       210       220       230       240
m121-1.pep  PAFGFDTGPGNMLMDAWTQAHWQLPYDKNGAKAAQGNILPQLLDRLLAHPYFAQPHPKST
            |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
a121-1      PAFGFDTGPGNMLMDAWMQAHWQLPYDKNGAKAAQGNILPQLLDRLLAHPYFAQPHPKST
                 190       200       210       220       230       240

                 250       260       270       280       290       300
m121-1.pep  GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVCDAVSHAAADARQMYICGGGIRNPV
            |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
a121-1      GRELFALNWLETYLDGGENRYDVLRTLSRFTAQTVFDAVSHAAADARQMYICGGGIRNPV
                 250       260       270       280       290       300

                 310       320       330       340       350       360
m121-1.pep  LMADLAECFGTRVSLHSTADLNLDPQWVEAAXFAWLAACWINRIPGSPHKATGASKPCIL
            ||||||||||||||||||:||||||||||| |||:||||:|||||||||||||||||||||
a121        LMADLAECFGTRVSLHSTAELNLDPQWVEAAAFAWMAACWVNRIPGSPHKATGASKPCIL
                 310       320       330       340       350       360

m121-1.pep  XAGYYYX
             ||||||
a121        GAGYYYX
```

128 and 128-1

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1010>:

```
m128.seq  (partial)
       1  ATGACTGACA ACGCACTGCT CCATTTGGGC GAAGAACCCC GTTTTGATCA
      51  AATCAAAACC GAAGACATCA AACCCGCCCT GCAAACCGCC ATCGCCGAAG
     101  CGCGCGAACA AATCGCCGCC ATCAAAGCCC AAACGCACAC CGGCTGGGCA
     151  AACACTGTCG AACCCCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
     201  GGGCGTGGTG TCGCACCTCA ACTGCGTCGC CGACACGCCC GAACTGCGCG
     251  CCGTCTATAA CGAACTGATG CCCGAAATCA CCGTCTTCTT CACCGAAATC
     301  GGACAAGACA TCGAGCTGTA CAACCGCTTC AAAACCATCA AAAATTCCCC
     351  CGAATTCGAC ACCCTCTCCC CCGCACAAAA AACCAAACTC AACCAC
       1  TACGCCAGCG AAAAACTGCG CGAAGCCAAA TACGCGTTCA GCGAAACCGA
      51  wGTCAAAAAA TAyTTCCCyG TCGGCAAwGT ATTAAACGGA CTGTTCGCCC
     101  AAmTCAAAAA ACTmTACGGC ATCGGATTTA CCGAAAAAAC yGTCCCCGTC
     151  TGGCACAAAG ACGTGCGCTA TTkTGAATTG CAACAAAACG GCGAAmCCAT
     201  AGGCGGCGTT TATATGGATT TGTACGCACG CGAAGGCAAA CGCGGCGGCG
     251  CGTGGATGAA CGACTACAAA GGCCGCCGCC GTTTTTCAGA CGGCACGCTG
     301  CAAyTGCCCA CCGCCTACCT CGTCTGCAAC TTCGCCCCAC CCGTCGGCGG
     351  CAGGGAAGCC CGCyTGAGCC ACGACGAAAT CCTCATCCTC TTCCACGAAA
     401  CCGGACACGG GCTGCACCAC CTGCTTACCC AAGTGGACGA ACTGGGCGTA
     451  TCCGGCATCA ACGGCGTAkA ATGGGACGCG GTCGAACTGC CCAGCCAGTT
     501  TATGGAAAAT TTCGTTTGGG AATACAATGT CTTGGCACAA mTGTCAGCCC
     551  ACGAAGAAAC CGGcgTTCCC yTGCCGAAAG AACTCTTsGA CAAAwTGCTC
     601  GCCGCCAAAA ACTTCCAAsG CGGCATGTTC yTsGTCCGGC AAwTGGAGTT
     651  CGCCCTCTTT GATATGATGA TTTACAGCGA AGACGACGAA GGCCGTCTGA
     701  AAAACTGGCA ACAGGTTTTA GACAGCGTGC GCAAAAAAGT CGCCGTCATC
     751  CAGCCGCCCG AATACAACCG CTTCGCCTTG AGCTTCGGCC ACATCTTCGC
     801  AGGCGGCTAT TCCGCAGCTn ATTACAGCTA CGCGTGGGCG GAAGTATTGA
```

```
 851  GCGCGGACGC ATACGCCGCC TTTGAAGAAA GCGACGATGT CGCCGCCACA
 901  GGCAAACGCT TTTGGCAGGA AATCCTCGCC GTCGGGGnAT CGCGCAGCGG
 951  nGCAGAATCC TTCAAAGCCT TCCGCGGCCG CGAACCGAGC ATAGACGCAC
1001  TCTTGCGCCA CAGCGGTTTC GACAACGCGG TCTGA
```

This corresponds to the amino acid sequence <SEQ ID 1011; ORF 128>:

```
m128.pep   (partial)
   1  MTDNALLHLG EEPRFDQIKT EDIKPALQTA IAEAREQIAA IKAQTHTGWA
  51  NTVEPLTGIT ERVGRIWGVV SHLNCVADTP ELRAVYNELM PEITVFFTEI
 101  GQDIELYNRF KTIKNSPEFD TLSPAQKTKL NH
//
   1  YASEKLREAK YAFSETXVKK YFPVGXVLNG LFAQXKKLYG IGFTEKTVPV
  51  WHKDVRYXEL QQNGEXIGGV YMDLYAREGK RGGAWMNDYK GRRRFSDGTL
 101  QLPTAYLVCN FAPPVGGREA RLSHDEILIL FHETGHGLHH LLTQVDELGV
 151  SGINGVXWDA VELPSQFMEN FVWEYNVLAQ XSAHEETGVP LPKELXDKXL
 201  AAKNFQXGMF XVRQXEFALF DMMIYSEDDE GRLKNWQQVL DSVRKKVAVI
 251  QPPEYNRFAL SFGHIFAGGY SAAXYSYAWA EVLSADAYAA FEESDDVAAT
 301  GKRFWQEILA VGXSRSGAES FKAFRGREPS IDALLRHSGF DNAV*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1012>:

```
g128.seq
   1  atgattgaca acgCActgct ccacttgggc gaagaacccC GTTTTaatca
  51  aatccaaacc gaagACAtca AACCCGCCGT CCAAACCGCC ATCGCCGAAG
 101  CGCGCGGACA AATCGCCGCC GTCAAAGCGC AAACGCACAC CGGCTGGGCG
 151  AACACCGTCG AGCGTCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
 201  GGGCGTCGTG TCCCATCTCA ACTCCGTCGT CGACACGCCC GAACTGCGCG
 251  CCGTCTATAA CGAACTGATG CCTGAAATCA CCGTCTTCTT CACCGAAATC
 301  GGACAAGACA TCGAACTGTA CAACCGCTTC AAAACCATCA AAAATTCCCC
 351  CGAATTTGCA ACGCTTTCCC CCGCACAAAA AACCAAGCTC GATCACGACC
 401  TGCGCGATTT CGTATTGAGC GGCGCGGAAC TGCCGCCCGA ACGGCAGGCA
 451  GAACTGGCAA AACTGCAAAC CGAAGGCGCG CAACTTTCCG CCAAATTCTC
 501  CCAAAACGTC CTAGACGCGA CCGACGCGTT CGGCATTTAC TTTGACGATG
 551  CCGCACCGCT TGCCGGCATT CCCGAAGACG CGCTCGCCAT GTTTGCCGCC
 601  GCCGCGCAAA GCGAAGGCAA AACAGGTTAC AAAATCGGCT TGCAGATTCC
 651  GCACTACCTT GCCGTTATCC AATACGCCGG CAACCGCGAA CTGCGCGAAC
 701  AAATCTACCG CGCCTACGTT ACCCGTGCCA GCGAACTTTC AAACGACGGC
 751  AAATTCGACA ACACCGCCAA CATCGACCGC ACGCTCGAAA ACGCATTGAA
 801  AACCGccaaa cTGCTCGGCT TTAAAAATTA CGCCGAATTG TCGCTGGCAA
 851  CCAAAATGGC GGACACGCCC GAACAGGTTT TAAACTTCCT GCACGACCTC
 901  GCCCGCCGCG CCAAACCCTA CGCCGAAAAA GACCTCGCCG AAGTCAAAGC
 951  CTTCGCCCGC GAACACCTCG GTCTCGCCGA CCCGCAGCCG TGGGACTTGA
1001  GCTACGCCGG CGAAAAACTG CGCGAAGCCA AATACGCATT CAGCGAAACC
1051  GAAGTCAAAA AATACTTCCC CGTCGGCAAA GTTCTGGCAG GCCTGTTCGC
1101  CCAAATCAAA AAACTCTACG GCATCGGATT CGCCGAAAAA ACCGTTCCCG
1151  TCTGGCACAA AGACGTGCGC TATTTTGAAT TGCAACAAAA CGGCAAAACC
1201  ATCGGCGGCG TTTATATGGA TTTGTACGCA CGCGAAGGCA AACGCGGCGG
1251  CGCGTGGATG AACGACtaca AAGGCCGCCG CCGCTTTGCC GACGgcacGC
1301  TGCAACTGCC CACCGCCTAC CTCGTCTGCA ACTTCGCCCC GCCCGTCGGC
1351  GGCAAAGAAG CGCGTTTAAG CCACGACGAA ATCCTCACCC TCTTCCACGA
1401  AacCGGCCAC GGACTGCACC ACCTGCTTAC CCAAGTGGAC GAACTGGGCG
1451  TGTCCGGCAT CAAcggcgtA GAATGGGACG CGGTCGAACT GCCCAGCCAG
1501  TTTATGGAAA ACTTCGTTTG GGAATACAAT GTATTGGCAC AAATGTCCGC
1551  CCACGAAGAA AccgGCGAGC CCCTGCCGAA AGAACTCTTC GACAAAATGC
1601  TcgcCGCCAA AAACTTCCAG CGCGGTATGT TCCTCGTCCG GCAAATGGAG
1651  TTCGCCCTCT TCGATATGAT GATTTACAGT GAAAGCGACG AATGCCGTCT
1701  GAAAAACTGG CAGCAGGTTT TAGACAGCGT GCGCAAAGAA GTcGCCGTCA
1751  TCCAACCGCC CGAATACAAC CGCTTCGCCA ACAGCTTCGG CCacatctTC
1801  GCcggcGGCT ATTCCGCAGG CTATTACAGC TACGCATGGG CCGAAGTCCt
1851  cAGCACCGAT GCCTACGCCG CCTTTGAAGA AAGcGACGac gtcGCCGCCA
```

```
1901  CAGGCAAACG CTTCTGGCAA GAAAtccttg ccgtcggcgg ctCCCGCAGC
1951  gcgGCGGAAT CCTTCAAAGC CTTCCGCGGA CGCGAACCGA GCATAGACGC
2001  ACTGCTGCGC CAaagcggtT TCGACAACGC gGCttgA
```

This corresponds to the amino acid sequence <SEQ ID 1013; ORF 128.ng>:

```
g128.pep
   1  MIDNALLHLG EEPRFNQIQT EDIKPAVQTA IAEARGQIAA VKAQTHTGWA
  51  NTVERLTGIT ERVGRIWGVV SHLNSVVDTP ELRAVYNELM PEITVFFTEI
 101  GQDIELYNRF KTIKNSPEFA TLSPAQKTKL DHDLRDFVLS GAELPPERQA
 151  ELAKLQTEGA QLSAKFSQNV LDATDAFGIY FDDAAPLAGI PEDALAMFAA
 201  AAQSEGKTGY KIGLQIPHYL AVIQYAGNRE LREQIYRAYV TRASELSNDG
 251  KFDNTANIDR TLENALKTAK LLGFKNYAEL SLATKMADTP EQVLNFLHDL
 301  ARRAKPYAEK DLAEVKAFAR EHLGLADPQP WDLSYAGEKL REAKYAFSET
 351  EVKKYFPVGK VLAGLFAQIK KLYGIGFAEK TVPVWHKDVR YFELQQNGKT
 401  IGGVYMDLYA REGKRGGAWM NDYKGRRRFA DGTLQLPTAY LVCNFAPPVG
 451  GKEARLSHDE ILTLFHETGH GLHHLLTQVD ELGVSGINGV EWDAVELPSQ
 501  FMENFVWEYN VLAQMSAHEE TGEPLPKELF DKMLAAKNFQ RGMFLVRQME
 551  FALFDMMIYS ESDECRLKNW QQVLDSVRKE VAVIQPPEYN RFANSFGHIF
 601  AGGYSAGYYS YAWAEVLSTD AYAAFEESDD VAATGKRFWQ EILAVGGSRS
 651  AAESFKAFRG REPSIDALLR QSGFDNAA*
```

ORF 128 shows 91.7% identity over a 475 aa overlap with a predicted ORF (ORF 128.ng) from *N. gonorrhoeae:*

```
m128/g128

                       10        20        30        40        50        60
g128.pep       MIDNALLHLGEEPRFNQIQTEDIKPAVQTAIAEARGQIAAVKAQTHTGWANTVERLTGIT
               | |||||||||||||:||:|||||||:|||||||| ||||:||||||||||||| |||||
m128           MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
                       10        20        30        40        50        60

                       70        80        90       100       110       120
g128.pep       ERVGRIWGVVSHLNSVVDTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFA
               |||||||||||||| |:|||||||||||||||||||||||||||||||||||||||||| 
m128           ERVGRIWGVVSHLNCVADTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
                       70        80        90       100       110       120

                      130       140       150       160       170       180
g128.pep       TLSPAQKTKLDHDLRDFVLSGAELPPERQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
               ||||||||||:|
m128           TLSPAQKTKLNH
                      130
                             //
                                                  340       350       360
g128.pep                                     YAGEKLREAKYAFSETEVKKYFPVGKVLAG
                                             ||:||||||||||||| |||||||| || |
m128                                         YASEKLREAKYAFSETXVKKYFPVGXVLNG
                                                       10        20        30

                  370       380       390       400       410       420
g128.pep       LFAQIKKLYGIGFAEKTVPVWHKDVRYFELQQNGKTIGGVYMDLYAREGKRGGAWMNDYK
               |||| ||||||||:||||||||||||| ||||||::||||||||||||||||||||||||
m128           LFAQXKKLYGIGFTEKTVPVWHKDVRYXELQQNGEXIGGVYMDLYAREGKRGGAWMNDYK
                       40        50        60        70        80        90

                  430       440       450       460       470       480
g128.pep       GRRRFADGTLQLPTAYLVCNFAPPVGGKEARLSHDEILTLFHETGHGLHHLLTQVDELGV
```

```
                |||||:|||||||||||||||||||||:|||||||||| |||||||||||||||||||||
m128            GRRRFSDGTLQLPTAYLVCNFAPPVGGREARLSHDEILILFHETGHGLHHLLTQVDELGV
                       100       110       120       130       140       150

                   490       500       510       520       530       540
g128.pep        SGINGVEWDAVELPSQFMENFVWEYNVLAQMSAHEETGEPLPKELFDKMLAAKNFQRGMF
                |||||| ||||||||||||||||||||||| |||||||| |||||| || ||||||| |||
m128            SGINGVXWDAVELPSQFMENFVWEYNVLAQXSAHEETGVPLPKELXDKXLAAKNFQXGMF
                       160       170       180       190       200       210

                   550       560       570       580       590       600
g128.pep        LVRQMEFALFDMMIYSESDECRLKNWQQVLDSVRKEVAVIQPPEYNRFANSFGHIFAGGY
                 ||| ||||||||||||:|| ||||||||||||||:||||||||||||| ||||||||||
m128            XVRQXEFALFDMMIYSEDDEGRLKNWQQVLDSVRKKVAVIQPPEYNRFALSFGHIFAGGY
                       220       230       240       250       260       270

                   610       620       630       640       650       660
g128.pep        SAGYYSYAWAEVLSTDAYAAFEESDDVAATGKRFWQEILAVGGSRSAAESFKAFRGREPS
                ||: ||||||||||:||||||||||||||||||||||||||| |||:|||||||||||||
m128            SAAXYSYAWAEVLSADAYAAFEESDDVAATGKRFWQEILAVGXSRSGAESFKAFRGREPS
                       280       290       300       310       320       330

                   670      679
g128.pep        IDALLRQSGFDNAAX
                ||||||:||||||:
m128            IDALLRHSGFDNAVX
                       340
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1014>:

```
a128.seq
       1  ATGACTGACA ACGCACTGCT CCATTTGGGC GAAGAACCCC GTTTTGATCA
      51  AATCAAAACC GAAGACATCA AACCCGCCCT GCAAACCGCC ATTGCCGAAG
     101  CGCGCGAACA AATCGCCGCC ATCAAAGCCC AAACGCACAC CGGCTGGGCA
     151  AACACTGTCG AACCCCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
     201  GGGCGTGGTG TCGCACCTCA ACTCCGTCAC CGACACGCCC GAACTGCGCG
     251  CCGCCTACAA TGAATTAATG CCCGAAATTA CCGTCTTCTT CACCGAAATC
     301  GGACAAGACA TCGAGCTGTA CAACCGCTTC AAAACCATCA AAAACTCCCC
     351  CGAGTTCGAC ACCCTCTCCC ACGCGCAAAA AACCAAACTC AACCACGATC
     401  TGCGCGATTT CGTCCTCAGC GGCGCGGAAC TGCCGCCCGA ACAGCAGGCA
     451  GAATTGGCAA AACTGCAAAC CGAAGGCGCG CAACTTTCCG CCAAATTCTC
     501  CCAAAACGTC CTAGACGCGA CCGACGCGTT CGGCATTTAC TTTGACGATG
     551  CCGCACCGCT TGCCGGCATT CCCGAAGACG CGCTCGCCAT GTTTGCCGCT
     601  GCCGCGCAAA GCGAAGGCAA AACAGGCTAC AAAATCGGTT TGCAGATTCC
     651  GCACTACCTC GCCGTCATCC AATACGCCGA CAACCGCAAA CTGCGCGAAC
     701  AAATCTACCG CGCCTACGTT ACCCGCGCCA GCGAGCTTTC AGACGACGGC
     751  AAATTCGACA ACACCGCCAA CATCGACCGC ACGCTCGAAA ACGCCCTGCA
     801  AACCGCCAAA CTGCTCGGCT TCAAAAACTA CGCCGAATTG TCGCTGGCAA
     851  CCAAAATGGC GGACACCCCC GAACAAGTTT TAAACTTCCT GCACGACCTC
     901  GCCCGCCGCG CCAAACCCTA CGCCGAAAAA GACCTCGCCG AAGTCAAAGC
     951  CTTCGCCCGC GAAAGCCTCG GCCTCGCCGA TTTGCAACCG TGGGACTTGG
    1001  GCTACGCCGG CGAAAAACTG CGCGAAGCCA AATACGCATT CAGCGAAACC
    1051  GAAGTCAAAA AATACTTCCC CGTCGGCAAA GTATTAAACG GACTGTTCGC
    1101  CCAAATCAAA AAACTCTACG GCATCGGATT TACCGAAAAA ACCGTCCCCG
    1151  TCTGGCACAA AGACGTGCGC TATTTTGAAT TGCAACAAAA CGGCGAAACC
    1201  ATAGGCGGCG TTTATATGGA TTTGTACGCA CGCGAAGGCA AACGCGGCGG
    1251  CGCGTGGATG AACGACTACA AAGGCCGCCG CCGTTTTTCA GACGGCACGC
    1301  TGCAACTGCC CACCGCCTAC CTCGTCTGCA ACTTCACCCC GCCCGTCGGC
    1351  GGCAAAGAAG CCCGCTTGAG CCATGACGAA ATCCTCACCC TCTTCCACGA
    1401  AACCGGACAC GGCCTGCACC ACCTGCTTAC CCAAGTCGAC GAACTGGGCG
    1451  TATCCGGCAT CAACGGCGTA GAATGGGACG CAGTCGAACT GCCCAGTCAG
    1501  TTTATGGAAA ATTTCGTTTG GGAATACAAT GTCTTGGCGC AAATGTCCGC
```

```
1551 CCACGAAGAA ACCGGCGTTC CCCTGCCGAA AGAACTCTTC GACAAAATGC
1601 TCGCCGCCAA AAACTTCCAA CGCGGAATGT TCCTCGTCCG CCAAATGGAG
1651 TTCGCCCTCT TTGATATGAT GATTTACAGC GAAGACGACG AAGGCCGTCT
1701 GAAAAACTGG CAACAGGTTT TAGACAGCGT GCGCAAAGAA GTCGCCGTCG
1751 TCCGACCGCC CGAATACAAC CGCTTCGCCA ACAGCTTCGG CCACATCTTC
1801 GCAGGCGGCT ATTCCGCAGG CTATTACAGC TACGCGTGGG CGGAAGTATT
1851 GAGCGCGGAC GCATACGCCG CCTTTGAAGA AAGCGACGAT GTCGCCGCCA
1901 CAGGCAAACG CTTTTGGCAG GAAATCCTCG CCGTCGGCGG ATCGCGCAGC
1951 GCGGCAGAAT CCTTCAAAGC CTTCCGCGGA CGCGAACCGA GCATAGACGC
2001 ACTCTTGCGC CACAGCGGCT TCGACAACGC GGCTTGA
```

This corresponds to the amino acid sequence <SEQ ID 1015; ORF 128.a>:

```
a128.pep
  1 MTDNALLHLG EEPRFDQIKT EDIKPALQTA IAEAREQIAA IKAQTHTGWA
 51 NTVEPLTGIT ERVGRIWGVV SHLNSVTDTP ELRAAYNELM PEITVFFTEI
101 GQDIELYNRF KTIKNSPEFD TLSHAQKTKL NHDLRDFVLS GAELPPEQQA
151 ELAKLQTEGA QLSAKFSQNV LDATDAFGIY FDDAAPLAGI PEDALAMFAA
201 AAQSEGKTGY KIGLQIPHYL AVIQYADNRK LREQIYRAYV TRASELSDDG
251 KFDNTANIDR TLENALQTAK LLGFKNYAEL SLATKMADTP EQVLNFLHDL
301 ARRAKPYAEK DLAEVKAFAR ESLGLADLQP WDLGYAGEKL REAKYAFSET
351 EVKKYFPVGK VLNGLFAQIK KLYGIGFTEK TVPVWHKDVR YFELQQNGET
401 IGGVYMDLYA REGKRGGAWM NDYKGRRRFS DGTLQLPTAY LVCNFTPPVG
451 GKEARLSHDE ILTLFHETGH GLHHLLTQVD ELGVSGINGV EWDAVELPSQ
501 FMENFVWEYN VLAQMSAHEE TGVPLPKELF DKMLAAKNFQ RGMFLVRQME
551 FALFDMMIYS EDDEGRLKNW QQVLDSVRKE VAVVRPPEYN RFANSFGHIF
601 AGGYSAGYYS YAWAEVLSAD AYAAFEESDD VAATGKRFWQ EILAVGGSRS
651 AAESFKAFRG REPSIDALLR HSGFDNAA*
```

m128/a128 ORFs 128 and 128.a showed a 66.0% identity in 677 aa overlap

```
                    10        20        30        40        50        60
m128.pep    MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a128        MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
                    10        20        30        40        50        60

                    70        80        90       100       110       120
m128.pep    ERVGRIWGVVSHLNCVADTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
            |||||||||||||| |:||||||:||||||||||||||||||||||||||||||||||||
a128        ERVGRIWGVVSHLNSVTDTPELRAAYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
                    70        80        90       100       110       120

                   130
m128.pep    TLSPAQKTKLNH------------------------------------------------
            ||| ||||||||
a128        TLSHAQKTKLNHDLRDFVLSGAELPPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
                   130       140       150       160       170       180

m128.pep    ------------------------------------------------------------

a128        FDDAAPLAGIPEDALAMFAAAAQSEGKTGYKIGLQIPHYLAVIQYADNRKLREQIYRAYV
                   190       200       210       220       230       240

m128.pep    ------------------------------------------------------------

a128        TRASELSDDGKFDNTANIDRTLENALQTAKLLGFKNYAELSLATKMADTPEQVLNFLHDL
                   250       260       270       280       290       300

                                                  140       150
m128.pep    ----------------------------------YASEKLREAKYAFSETXVKKYFPVGX
                                              ||:||||||||||||| ||||||||
```

```
a128          ARRAKPYAEKDLAEVKAFARESLGLADLQPWDLGYAGEKLREAKYAFSETEVKKYFPVGK
                      310       320       330       340       350       360

              160       170       180       190       200       210
m128.pep      VLNGLFAQXKKLYGIGFTEKTVPVWHKDVRYXELQQNGEXIGGVYMDLYAREGKRGGAWM
              |||||||| |||||||||||||||||||||| |||||||:||||||||||||||||||||
a128          VLNGLFAQIKKLYGIGFTEKTVPVWHKDVRYFELQQNGETIGGVYMDLYAREGKRGGAWM
                      370       380       390       400       410       420

              220       230       240       250       260       270
m128.pep      NDYKGRRRFSDGTLQLPTAYLVCNFAPPVGGREARLSHDEILILFHETGHGLHHLLTQVD
              |||||||||||||||||||||||||:|||||:|||||||||| |||||||||||||||||
a128          NDYKGRRRFSDGTLQLPTAYLVCNFTPPVGGKEARLSHDEILTLFHETGHGLHHLLTQVD
                      430       440       450       460       470       480

              280       290       300       310       320       330
m128.pep      ELGVSGINGVXWDAVELPSQFMENFVWEYNVLAQXSAHEETGVPLPKELXDKXLAAKNFQ
              |||||||||| ||||||||||||||||||||||| |||||||||||||| || |||||||
a128          ELGVSGINGVEWDAVELPSQFMENFVWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFQ
                      490       500       510       520       530       540

              340       350       360       370       380       390
m128.pep      XGMFXVRQXEFALFDMMIYSEDDEGRLKNWQQVLDSVRKKVAVIQPPEYNRFALSFGHIF
               ||| ||| ||||||||||||||||||||||||||||||:|||::|||||||| ||||||
a128          RGMFLVRQMEFALFDMMIYSEDDEGRLKNWQQVLDSVRKEVAVVRPPEYNRFANSFGHIF
                      550       560       570       580       590       600

              400       410       420       430       440       450
m128.pep      AGGYSAAXYSYAWAEVLSADAYAAFEESDDVAATGKRFWQEILAVGXSRSGAESFKAFRG
              ||||||: |||||||||||||||||||||||||||||||||||||| |||:|||||||||
a128          AGGYSAGYYSYAWAEVLSADAYAAFEESDDVAATGKRFWQEILAVGGSRSAAESFKAFRG
                      610       620       630       640       650       660

              460       470
m128.pep      REPSIDALLRHSGFDNAVX
              |||||||||||||||||:
a128          REPSIDALLRHSGFDNAAX
                      670
```

Further work revealed the DNA sequence identified in *N. meningitidis* <SEQ ID 1016>:

```
m128-1.seq
     1  ATGACTGACA ACGCACTGCT CCATTTGGGC GAAGAACCCC GTTTTGATCA
    51  AATCAAAACC GAAGACATCA AACCCGCCCT GCAAACCGCC ATCGCCGAAG
   101  CGCGCGAACA AATCGCCGCC ATCAAAGCCC AAACGCACAC CGGCTGGGCA
   151  AACACTGTCG AACCCCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
   201  GGGCGTGGTG TCGCACCTCA ACTCCGTCGC CGACACGCCC GAACTGCGCG
   251  CCGTCTATAA CGAACTGATG CCCGAAATCA CCGTCTTCTT CACCGAAATC
   301  GGACAAGACA TCGAGCTGTA CAACCGCTTC AAAACCATCA AAAATTCCCC
   351  CGAATTCGAC ACCCTCTCCC CCGCACAAAA AACCAAACTC AACCACGATC
   401  TGCGCGATTT CGTCCTCAGC GGCGCGGAAC TGCCGCCCGA ACAGCAGGCA
   451  GAACTGGCAA AACTGCAAAC CGAAGGCGCG CAACTTTCCG CCAAATTCTC
   501  CCAAAACGTC CTAGACGCGA CCGACGCGTT CGGCATTTAC TTTGACGATG
   551  CCGCACCGCT TGCCGGCATT CCCGAAGACG CGCTCGCCAT GTTTGCCGCC
   601  GCCGCGCAAA GCGAAAGCAA AACAGGCTAC AAAATCGGCT TGCAGATTCC
   651  ACACTACCTC GCCGTCATCC AATACGCCGA CAACCGCGAA CTGCGCGAAC
   701  AAATCTACCG CGCCTACGTT ACCCGCGCCA GCGAACTTTC AGACGACGGC
   751  AAATTCGACA ACACCGCCAA CATCGACCGC ACGCTCGCAA ACGCCCTGCA
   801  AACCGCCAAA CTGCTCGGCT TCAAAAACTA CGCCGAATTG TCGCTGGCAA
   851  CCAAAATGGC GGACACGCCC GAACAAGTTT TAAACTTCCT GCACGACCTC
```

```
 901  GCCCGCCGCG CCAAACCCTA CGCCGAAAAA GACCTCGCCG AAGTCAAAGC
 951  CTTCGCCCGC GAAAGCCTGA ACCTCGCCGA TTTGCAACCG TGGGACTTGG
1001  GCTACGCCAG CGAAAAACTG CGCGAAGCCA AATACGCGTT CAGCGAAACC
1051  GAAGTCAAAA AATACTTCCC CGTCGGCAAA GTATTAAACG GACTGTTCGC
1101  CCAAATCAAA AAACTCTACG GCATCGGATT TACCGAAAAA ACCGTCCCCG
1151  TCTGGCACAA AGACGTGCGC TATTTTGAAT TGCAACAAAA CGGCGAAACC
1201  ATAGGCGGCG TTTATATGGA TTTGTACGCA CGCGAAGGCA AACGCGGCGG
1251  CGCGTGGATG AACGACTACA AAGGCCGCCG CCGTTTTTCA GACGGCACGC
1301  TGCAACTGCC CACCGCCTAC CTCGTCTGCA ACTTCGCCCC ACCCGTCGGC
1351  GGCAGGGAAG CCCGCCTGAG CCACGACGAA ATCCTCATCC TCTTCCACGA
1401  AACCGGACAC GGGCTGCACC ACCTGCTTAC CCAAGTGGAC GAACTGGGCG
1451  TATCCGGCAT CAACGGCGTA GAATGGGACG CGGTCGAACT GCCCAGCCAG
1501  TTTATGGAAA ATTTCGTTTG GGAATACAAT GTCTTGGCAC AAATGTCAGC
1551  CCACGAAGAA ACCGGCGTTC CCCTGCCGAA AGAACTCTTC GACAAAATGC
1601  TCGCCGCCAA AAACTTCCAA CGCGGCATGT TCCTCGTCCG GCAAATGGAG
1651  TTCGCCCTCT TTGATATGAT GATTTACAGC GAAGACGACG AAGGCCGTCT
1701  GAAAAACTGG CAACAGGTTT TAGACAGCGT GCGCAAAAAA GTCGCCGTCA
1751  TCCAGCCGCC CGAATACAAC CGCTTCGCCT TGAGCTTCGG CCACATCTTC
1801  GCAGGCGGCT ATTCCGCAGG CTATTACAGC TACGCGTGGG CGGAAGTATT
1851  GAGCGCGGAC GCATACGCCG CCTTTGAAGA AAGCGACGAT GTCGCCGCCA
1901  CAGGCAAACG CTTTTGGCAG GAAATCCTCG CCGTCGGCGG ATCGCGCAGC
1951  GCGGCAGAAT CCTTCAAAGC CTTCCGCGGC CGCGAACCGA GCATAGACGC
2001  ACTCTTGCGC CACAGCGGTT TCGACAACGC GGTCTGA
```

This corresponds to the amino acid sequence <SEQ ID 1017; ORF 128-1>:

```
m128-1.pep.
  1  MTDNALLHLG EEPRFDQIKT EDIKPALQTA IAEAREQIAA IKAQTHTGWA
 51  NTVEPLTGIT ERVGRIWGVV SHLNSVADTP ELRAVYNELM PEITVFFTEI
101  GQDIELYNRF KTIKNSPEFD TLSPAQKTKL NHDLRDFVLS GAELPPEQQA
151  ELAKLQTEGA QLSAKFSQNV LDATDAFGIY FDDAAPLAGI PEDALAMFAA
201  AAQSESKTGY KIGLQIPHYL AVIQYADNRE LREQIYRAYV TRASELSDDG
251  KFDNTANIDR TLANALQTAK LLGFKNYAEL SLATKMADTP EQVLNFLHDL
301  ARRAKPYAEK DLAEVKAFAR ESLNLADLQP WDLGYASEKL REAKYAFSET
351  EVKKYFPVGK VLNGLFAQIK KLYGIGFTEK TVPVWHKDVR YFELQQNGET
401  IGGVYMDLYA REGKRGGAWM NDYKGRRRFS DGTLQLPTAY LVCNFAPPVG
451  GREARLSHDE ILILFHETGH GLHHLLTQVD ELGVSGINGV EWDAVELPSQ
501  FMENFVWEYN VLAQMSAHEE TGVPLPKELF DKMLAAKNFQ RGMFLVRQME
551  FALFDMMIYS EDDEGRLKNW QQVLDSVRKK VAVIQPPEYN RFALSFGHIF
601  AGGYSAGYYS YAWAEVLSAD AYAAFEESDD VAATGKRFWQ EILAVGGSRS
651  AAESFKAFRG REPSIDALLR HSGFDNAV*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1018>:

```
g128-1.seq  (partial)
  1  ATGATTGACA ACGCACTGCT CCACTTGGGC GAAGAACCCC GTTTTAATCA
 51  AATCAAAACC GAAGACATCA AACCCGCCGT CCAAACCGCC ATCGCCGAAG
101  CGCGCGGACA AATCGCCGCC GTCAAAGCGC AAACGCACAC CGGCTGGGCG
151  AACACCGTCG AGCGTCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
201  GGGCGTCGTG TCCCATCTCA ACTCCGTCGT CGACACGCCC GAACTGCGCG
251  CCGTCTATAA CGAACTGATG CCTGAAATCA CCGTCTTCTT CACCGAAATC
301  GGACAAGACA TCGAACTGTA CAACCGCTTC AAAACCATCA AAAATTCCCC
351  CGAATTTGCA ACGCTTTCCC CCGCACAAAA AACCAAGCTC GATCACGACC
401  TGCGCGATTT CGTATTGAGC GGCGCGGAAC TGCCGCCCGA ACGGCAGGCA
451  GAACTGGCAA AACTGCAAAC CGAAGGCGCG CAACTTTCCG CCAAATTCTC
501  CCAAAACGTC CTAGACGCGA CCGACGCGTT CGGCATTTAC TTTGACGATG
551  CCGCACCGCT TGCCGGCATT CCCGAAGACG CGCTCGCCAT GTTTGCCGCC
601  GCCGCGCAAA GCGAAGGCAA AACAGGTTAC AAAATCGGCT TGCAGATTCC
651  GCACTACCTT GCCGTTATCC AATACGCCGG CAACCGCGAA CTGCGCGAAC
701  AAATCTACCG CGCCTACGTT ACCCGTGCCA GCGAACTTTC AAACGACGGC
751  AAATTCGACA ACACCGCCAA CATCGACCGC ACGCTCGAAA ACGCATTGAA
801  AACCGCCAAA CTGCTCGGCT TTAAAAATTA CGCCGAATTG TCGCTGGCAA
851  CCAAAATGGC GGACACGCCC GAACAGGTTT TAAACTTCCT GCACGACCTC
901  GCCCGCCGCG CCAAACCCTA CGCCGAAAAA GACCTCGCCG AAGTCAAAGC
```

```
 951  CTTCGCCCGC GAACACCTCG GTCTCGCCGA CCCGCAGCCG TGGGACTTGA
1001  GCTACGCCGG CGAAAAACTG CGCGAAGCCA AATACGCATT CAGCGAAACC
1051  GAAGTCAAAA AATACTTCCC CGTCGGCAAA GTTCTGGCAG GCCTGTTCGC
1101  CCAAATCAAA AAACTCTACG GCATCGGATT CGCCGAAAAA ACCGTTCCCG
1151  TCTGGCACAA AGACGTGCGC TATTTTGAAT TGCAACAAAA CGGCAAAACC
1201  ATCGGCGGCG TTTATATGGA TTTGTACGCA CGCGAAGGCA AACGCGGCGG
1251  CGCGTGGATG AACGACTACA AAGGCCGCCG CCGCTTTGCC GACGGCACGC
1301  TGCAACTGCC CACCGCCTAC CTCGTCTGCA ACTTCGCCCC GCCCGTCGGC
1351  GGCAAAGAAG CGCGTTTAAG CCACGACGAA ATCCTCACCC TCTTCCACGA
1401  AACCGGCCAC GGACTGCACC ACCTGCTTAC CCAAGTGGAC GAACTGGGCG
1451  TGTCCGGCAT CAACGGCGTA AAA
```

This corresponds to the amino acid sequence <SEQ ID 1019; ORF 128-1.ng>:

```
g128-1.pep  (partial)
  1  MIDNALLHLG EEPRFNQIKT EDIKPAVQTA IAEARGQIAA VKAQTHTGWA
 51  NTVERLTGIT ERVGRIWGVV SHLNSVVDTP ELRAVYNELM PEITVFFTEI
101  GQDIELYNRF KTIKNSPEFA TLSPAQKTKL DHDLRDFVLS GAELPPERQA
151  ELAKLQTEGA QLSAKFSQNV LDATDAFGIY FDDAAPLAGI PEDALAMFAA
201  AAQSEGKTGY KIGLQIPHYL AVIQYAGNRE LREQIYRAYV TRASELSNDG
251  KFDNTANIDR TLENALKTAK LLGFKNYAEL SLATKMADTP EQVLNFLHDL
301  ARRAKPYAEK DLAEVKAFAR EHLGLADPQP WDLSYAGEKL REAKYAFSET
351  EVKKYFPVGK VLAGLFAQIK KLYGIGFAEK TVPVWHKDVR YFELQQNGKT
401  IGGVYMDLYA REGKRGGAWM NDYKGRRRFA DGTLQLPTAY LVCNFAPPVG
451  GKEARLSHDE ILTLFHETGH GLHHLLTQVD ELGVSGINGV K
```

m128-1/g128-1 ORFs 128-1 and 128-1.ng showed a 94.5% identity in 491 aa overlap

```
                      10        20        30        40        50        60
g128-1.pep  MIDNALLHLGEEPRFNQIKTEDIKPAVQTAIAEARGQIAAVKAQTHTGWANTVERLTGIT
            | |||||||||||||:||||||||||:|||||||| ||||:|||||||||||||| ||||
m128-1      MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
                      10        20        30        40        50        60

                      70        80        90       100       110       120
g128-1.pep  ERVGRIWGVVSHLNSVVDTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFA
            ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||| 
m128-1      ERVGRIWGVVSHLNSVADTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
                      70        80        90       100       110       120

                     130       140       150       160       170       180
g128-1.pep  TLSPAQKTKLDHDLRDFVLSGAELPPERQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
            ||||||||||:||||||||||||||||:||||||||||||||||||||||||||||||||
m128-1      TLSPAQKTKLNHDLRDFVLSGAELPPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
                     130       140       150       160       170       180

                     190       200       210       220       230       240
g128-1.pep  FDDAAPLAGIPEDALAMFAAAAQSEGKTGYKIGLQIPHYLAVIQYAGNRELREQIYRAYV
            |||||||||||||||||||||||||:|||||||||||||||||||| |||||||||||||
m128-1      FDDAAPLAGIPEDALAMFAAAAQSESKTGYKIGLQIPHYLAVIQYADNRELREQIYRAYV
                     190       200       210       220       230       240

                     250       260       270       280       290       300
g128-1.pep  TRASELSNDGKFDNTANIDRTLENALKTAKLLGFKNYAELSLATKMADTPEQVLNFLHDL
            |||||||:||||||||||||||| ||:|||||||||||||||||||||||||||||||||
m128-1      TRASELSDDGKFDNTANIDRTLANALQTAKLLGFKNYAELSLATKMADTPEQVLNFLHDL
                     250       260       270       280       290       300

                     310       320       330       340       350       360
g128-1.pep  ARRAKPYAEKDLAEVKAFAREHLGLADPQPWDLSYAGEKLREAKYAFSETEVKKYFPVGK
            |||||||||||||||||||| |:||| |||||:||:||||||||||||||||||||||||
```

```
m128-1        ARRAKPYAEKDLAEVKAFARESLNLADLQPWDLGYASEKLREAKYAFSETEVKKYFPVGK
                     310       320       330       340       350       360

                     370       380       390       400       410       420
g128-1.pep    VLAGLFAQIKKLYGIGFAEKTVPVWHKDVRYFELQQNGKTIGGVYMDLYAREGKRGGAWM
              || ||||||||||||||:|||||||||||||||||||||:||||||||||||||||||||
m128-1        VLNGLFAQIKKLYGIGFTEKTVPVWHKDVRYFELQQNGETIGGVYMDLYAREGKRGGAWM
                     370       380       390       400       410       420

                     430       440       450       460       470       480
g128-1.pep    NDYKGRRRFADGTLQLPTAYLVCNFAPPVGGKEARLSHDEILTLFHETGHGLHHLLTQVD
              |||||||||:|||||||||||||||||||||:|||||||||| |||||||||||||||||
m128-1        NDYKGRRRFSDGTLQLPTAYLVCNFAPPVGGREARLSHDEILILFHETGHGLHHLLTQVD
                     430       440       450       460       470       480

                     490
g128-1.pep    ELGVSGINGVK
              ||||||||||:
m128-1        ELGVSGINGVEWDAVELPSQFMENFVWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFQ
                     490       500       510       520       530       540
```

The following DNA sequence was identified in *N. meningitidis* <SEQ ID 1020>:

```
a128-1.seq
      1  ATGACTGACA ACGCACTGCT CCATTTGGGC GAAGAACCCC GTTTTGATCA
     51  AATCAAAACC GAAGACATCA AACCCGCCCT GCAAACCGCC ATTGCCGAAG
    101  CGCGCGAACA AATCGCCGCC ATCAAAGCCC AAACGCACAC CGGCTGGGCA
    151  AACACTGTCG AACCCCTGAC CGGCATCACC GAACGCGTCG GCAGGATTTG
    201  GGGCGTGGTG TCGCACCTCA ACTCCGTCAC CGACACGCCC GAACTGCGCG
    251  CCGCCTACAA TGAATTAATG CCCGAAATTA CCGTCTTCTT CACCGAAATC
    301  GGACAAGACA TCGAGCTGTA CAACCGCTTC AAAACCATCA AAAACTCCCC
    351  CGAGTTCGAC ACCCTCTCCC ACGCGCAAAA AACCAAACTC AACCACGATC
    401  TGCGCGATTT CGTCCTCAGC GGCGCGGAAC TGCCGCCCGA ACAGCAGGCA
    451  GAATTGGCAA AACTGCAAAC CGAAGGCGCG CAACTTTCCG CCAAATTCTC
    501  CCAAAACGTC CTAGACGCGA CCGACGCGTT CGGCATTTAC TTTGACGATG
    551  CCGCACCGCT TGCCGGCATT CCCGAAGACG CGCTCGCCAT GTTTGCCGCT
    601  GCCGCGCAAA GCGAAGGCAA AACAGGCTAC AAAATCGGTT TGCAGATTCC
    651  GCACTACCTC GCCGTCATCC AATACGCCGA CAACCGCAAA CTGCGCGAAC
    701  AAATCTACCG CGCCTACGTT ACCCGCGCCA GCGAGCTTTC AGACGACGGC
    751  AAATTCGACA ACACCGCCAA CATCGACCGC ACGCTCGAAA ACGCCCTGCA
    801  AACCGCCAAA CTGCTCGGCT TCAAAAACTA CGCCGAATTG TCGCTGGCAA
    851  CCAAAATGGC GGACACCCCC GAACAAGTTT TAAACTTCCT GCACGACCTC
    901  GCCCGCCGCG CCAAACCCTA CGCCGAAAAA GACCTCGCCG AAGTCAAAGC
    951  CTTCGCCCGC GAAAGCCTCG GCCTCGCCGA TTTGCAACCG TGGGACTTGG
   1001  GCTACGCCGG CGAAAAACTG CGCGAAGCCA AATACGCATT CAGCGAAACC
   1051  GAAGTCAAAA AATACTTCCC CGTCGGCAAA GTATTAAACG GACTGTTCGC
   1101  CCAAATCAAA AAACTCTACG GCATCGGATT TACCGAAAAA ACCGTCCCCG
   1151  TCTGGCACAA AGACGTGCGC TATTTTGAAT TGCAACAAAA CGGCGAAACC
   1201  ATAGGCGGCG TTTATATGGA TTTGTACGCA CGCGAAGGCA AACGCGGCGG
   1251  CGCGTGGATG AACGACTACA AAGGCCGCCG CCGTTTTTCA GACGGCACGC
   1301  TGCAACTGCC CACCGCCTAC CTCGTCTGCA ACTTCACCCC GCCCGTCGGC
   1351  GGCAAAGAAG CCCGCTTGAG CCATGACGAA ATCCTCACCC TCTTCCACGA
   1401  AACCGGACAC GGCCTGCACC ACCTGCTTAC CCAAGTCGAC GAACTGGGCG
   1451  TATCCGGCAT CAACGGCGTA GAATGGGACG CAGTCGAACT GCCCAGTCAG
   1501  TTTATGGAAA ATTTCGTTTG GGAATACAAT GTCTTGGCGC AAATGTCCGC
   1551  CCACGAAGAA ACCGGCGTTC CCCTGCCGAA AGAACTCTTC GACAAAATGC
   1601  TCGCCGCCAA AAACTTCCAA CGCGGAATGT TCCTCGTCCG CCAAATGGAG
   1651  TTCGCCCTCT TTGATATGAT GATTTACAGC GAAGACGACG AAGGCCGTCT
   1701  GAAAAACTGG CAACAGGTTT TAGACAGCGT GCGCAAAGAA GTCGCCGTCG
   1751  TCCGACCGCC CGAATACAAC CGCTTCGCCA ACAGCTTCGG CCACATCTTC
   1801  GCAGGCGGCT ATTCCGCAGG CTATTACAGC TACGCGTGGG CGGAAGTATT
   1851  GAGCGCGGAC GCATACGCCG CCTTTGAAGA AAGCGACGAT GTCGCCGCCA
   1901  CAGGCAAACG CTTTTGGCAG GAAATCCTCG CCGTCGGCGG ATCGCGCAGC
```

```
1951  GCGGCAGAAT CCTTCAAAGC CTTCCGCGGA CGCGAACCGA GCATAGACGC
2001  ACTCTTGCGC CACAGCGGCT TCGACAACGC GGCTTGA
```

This corresponds to the amino acid sequence <SEQ ID 1021; ORF 128-1.a>:

```
a128-1.pep
   1  MTDNALLHLG EEPRFDQIKT EDIKPALQTA IAEAREQIAA IKAQTHTGWA
  51  NTVEPLTGIT ERVGRIWGVV SHLNSVTDTP ELRAAYNELM PEITVFFTEI
 101  GQDIELYNRF KTIKNSPEFD TLSHAQKTKL NHDLRDFVLS GAELPPEQQA
 151  ELAKLQTEGA QLSAKFSQNV LDATDAFGIY FDDAAPLAGI PEDALAMFAA
 201  AAQSEGKTGY KIGLQIPHYL AVIQYADNRK LREQIYRAYV TRASELSDDG
 251  KFDNTANIDR TLENALQTAK LLGFKNYAEL SLATKMADTP EQVLNFLHDL
 301  ARRAKPYAEK DLAEVKAFAR ESLGLADLQP WDLGYAGEKL REAKYAFSET
 351  EVKKYFPVGK VLNGLFAQIK KLYGIGFTEK TVPVWHKDVR YFELQQNGET
 401  IGGVYMDLYA REGKRGGAWM NDYKGRRRFS DGTLQLPTAY LVCNFTPPVG
 451  GKEARLSHDE ILTLFHETGH GLHHLLTQVD ELGVSGINGV EWDAVELPSQ
 501  FMENFVWEYN VLAQMSAHEE TGVPLPKELF DKMLAAKNFQ RGMFLVRQME
 551  FALFDMMIYS EDDEGRLKNW QQVLDSVRKE VAVVRPPEYN RFANSFGHIF
 601  AGGYSAGYYS YAWAEVLSAD AYAAFEESDD VAATGKRFWQ EILAVGGSRS
 651  AAESFKAFRG REPSIDALLR HSGFDNAA*
```

m128-1/a128-1 ORFs 128-1 and 128-1.a showed a 97.8% identity in 677 aa overlap

```
                        10        20        30        40        50        60
a128-1.pep      MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m128-1          MTDNALLHLGEEPRFDQIKTEDIKPALQTAIAEAREQIAAIKAQTHTGWANTVEPLTGIT
                        10        20        30        40        50        60

                        70        80        90       100       110       120
a128-1.pep      ERVGRIWGVVSHLNSVTDTPELRAAYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
                ||||||||||||||||:|||||||:|||||||||||||||||||||||||||||||||||
m128-1          ERVGRIWGVVSHLNSVADTPELRAVYNELMPEITVFFTEIGQDIELYNRFKTIKNSPEFD
                        70        80        90       100       110       120

                       130       140       150       160       170       180
a128-1.pep      TLSHAQKTKLNHDLRDFVLSGAELPPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
                ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m128-1          TLSPAQKTKLNHDLRDFVLSGAELPPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIY
                       130       140       150       160       170       180

                       190       200       210       220       230       240
a128-1.pep      FDDAAPLAGIPEDALAMFAAAAQSEGKTGYKIGLQIPHYLAVIQYADNRKLREQIYRAYV
                |||||||||||||||||||||||||:|||||||||||||||||||||||:||||||||||
m128-1          FDDAAPLAGIPEDALAMFAAAAQSESKTGYKIGLQIPHYLAVIQYADNRELREQIYRAYV
                       190       200       210       220       230       240

                       250       260       270       280       290       300
a128-1.pep      TRASELSDDGKFDNTANIDRTLENALQTAKLLGFKNYAELSLATKMADTPEQVLNFLHDL
                |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
m128-1          TRASELSDDGKFDNTANIDRTLANALQTAKLLGFKNYAELSLATKMADTPEQVLNFLHDL
                       250       260       270       280       290       300

                       310       320       330       340       350       360
a128-1.pep      ARRAKPYAEKDLAEVKAFARESLGLADLQPWDLGYAGEKLREAKYAFSETEVKKYFPVGK
                |||||||||||||||||||||||:||||||||||||:|||||||||||||||||||||||
m128-1          ARRAKPYAEKDLAEVKAFARESLNLADLQPWDLGYASEKLREAKYAFSETEVKKYFPVGK
                       310       320       330       340       350       360

                       370       380       390       400       410       420
a128-1.pep      VLNGLFAQIKKLYGIGFTEKTVPVWHKDVRYFELQQNGETIGGVYMDLYAREGKRGGAWM
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m128-1          VLNGLFAQIKKLYGIGFTEKTVPVWHKDVRYFELQQNGETIGGVYMDLYAREGKRGGAWM
```

```
                 370       380       390       400       410       420

                 430       440       450       460       470       480
a128-1.pep  NDYKGRRRFSDGTLQLPTAYLVCNFTPPVGGKEARLSHDEILTLFHETGHGLHHLLTQVD
            |||||||||||||||||||||||||:|||||:|||||||||| |||||||||||||||||
m128-1      NDYKGRRRFSDGTLQLPTAYLVCNFAPPVGGREARLSHDEILILFHETGHGLHHLLTQVD
                 430       440       450       460       470       480

                 490       500       510       520       530       540
a128-1.pep  ELGVSGINGVEWDAVELPSQFMENFVWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFQ
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m128-1      ELGVSGINGVEWDAVELPSQFMENFVWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFQ
                 490       500       510       520       530       540

                 550       560       570       580       590       600
a128-1.pep  RGMFLVRQMEFALFDMMIYSEDDEGRLKNWQQVLDSVRKEVAVVRPPEYNRFANSFGHIF
            |||||||||||||||||||||||||||||||||||||||:|||::|||||||| ||||||
m128-1      RGMFLVRQMEFALFDMMIYSEDDEGRLKNWQQVLDSVRKKVAVIQPPEYNRFALSFGHIF
                 550       560       570       580       590       600

                 610       620       630       640       650       660
a128-1.pep  AGGYSAGYYSYAWAEVLSADAYAAFEESDDVAATGKRFWQEILAVGGSRSAAESFKAFRG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m128-1      AGGYSAGYYSYAWAEVLSADAYAAFEESDDVAATGKRFWQEILAVGGSRSAAESFKAFRG
                 610       620       630       640       650       660

                 670      679
a128-1.pep  REPSIDALLRHSGFDNAAX
            |||||||||||||||||:
m128-1      REPSIDALLRHSGFDNAVX
                 670
```

206

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1022>:

```
m206.seq
      1  ATGTTTCCCC CCGACAAAAC CCTTTTCCTC TGTCTCAGCG CACTGCTCCT
     51  CGCCTCATGC GGCACGACCT CCGGCAAACA CCGCCAACCG AAACCCAAAC
    101  AGACAGTCCG GCAAATCCAA GCCGTCCGCA TCAGCCACAT CGACCGCACA
    151  CAAGGCTCGC AGGAACTCAT GCTCCACAGC CTCGGACTCA TCGGCACGCC
    201  CTACAAATGG GGCGGCAGCA GCACCGCAAC CGGCTTCGAT TGCAGCGGCA
    251  TGATTCAATT CGTTTACAAr AACGCCCTCA ACGTCAAGCT GCCGCGCACC
    301  GCCCGCGACA TGGCGGCGGC AAGCCGGsAAA ATCCCCGAcA GCCGCyTCAA
    351  GGCCGGCGAC CTCGTATTCT TCAACACCGG CGGCGCACAC CGCTACTCAC
    401  ACGTCGGACT CTACATCGGC AACGGCGAAT TCATCCATGC CCCCAGCAGC
    451  GGCAAAACCA TCAAAACCGA AAAACTCTCC ACACCGTTTT ACGCCAAAAA
    501  CTACCTCGGC GCACATACTT TTTTTACAGA ATGA
```

This corresponds to the amino acid sequence <SEQ ID 1023; ORF 206>:

```
m206.pep..
      1  MFPPDKTLFL CLSALLLASC GTTSGKHRQP KPKQTVRQIQ AVRISHIDRT
     51  QGSQELMLHS LGLIGTPYKW GGSSTATGFD CSGMIQFVYK NALNVKLPRT
    101  ARDMAAASRK IPDSRXKAGD LVFFNTGGAH RYSHVGLYIG NGEFIHAPSS
    151  GKTIKTEKLS TPFYAKNYLG AHTFFTE*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1024>:

```
g206.seq
      1  atgttttccc ccgacaaaac ccttttcctc tgtctcggcg cactgctcct
```

```
 51 cgcctcatgc ggcacgacct ccggcaaaca ccgccaaccg aaacccaaac
101 agacagtccg gcaaatccaa gccgtccgca tcagccacat cggccgcaca
151 caaggctcgc aggaactcat gctccacagc ctcggactca tcggcacgcc
201 ctacaaatgg ggcggcagca gcaccgcaac cggcttcgac tgcagcggca
251 tgattcaatt ggtttacaaa aacgccctca acgtcaagct gccgcgcacc
301 gcccgcgaca tggcggcggc aagccgcaaa atccccgaca gccgcctcaa
351 ggccggcgac atcgtattct tcaacaccgg cggcgcacac cgctactcac
401 acgtcggact ctacatcggc aacggcgaat tcatccatgc cccggcagc
451 ggcaaaacca tcaaaaccga aaaactctcc acaccgtttt acgccaaaaa
501 ctaccttgga gcgcatacgt tttttacaga atga
```

This corresponds to the amino acid sequence <SEQ ID 1025; ORF 206.ng>:

```
g206.pep
  1 MFSPDKTLFL CLGALLLASC GTTSGKHRQP KPKQTVRQIQ AVRISHIGRT
 51 QGSQELMLHS LGLIGTPYKW GGSSTATGFD CSGMIQLVYK NALNVKLPRT
101 ARDMAAASRK IPDSRLKAGD IVFFNTGGAH RYSHVGLYIG NGEFIHAPGS
151 GKTIKTEKLS TPFYAKNYLG AHTFFTE*
```

ORF 206 shows 96.0% identity over a 177 aa overlap with a predicted ORF (ORF 206.ng) from *N. gonorrhoeae:*

```
m206/g206

                     10        20        30        40        50        60
m206.pep    MFPPDKTLFLCLSALLLASCGTTSGKHRQPKPKQTVRQIQAVRISHIDRTQGSQELMLHS
            || |||||||||||:||||||||||||||||||||||||||||||||||||| |||||||
g206        MFSPDKTLFLCLGALLLASCGTTSGKHRQPKPKQTVRQIQAVRISHIGRTQGSQELMLHS
                     10        20        30        40        50        60

                     70        80        90       100       110       120
m206.pep    LGLIGTPYKWGGSSTATGFDCSGMIQFVYKNALNVKLPRTARDMAAASRKIPDSRXKAGD
            |||||||||||||||||||||||||||||:||||||||||||||||||||||||||| ||||
g206        LGLIGTPYKWGGSSTATGFDCSGMIQLVYKNALNVKLPRTARDMAAASRKIPDSRLKAGD
                     70        80        90       100       110       120

                    130       140       150       160       170
m206.pep    LVFFNTGGAHRYSHVGLYIGNGEFIHAPSSGKTIKTEKLSTPFYAKNYLGAHTFFTEX
            :||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
g206        IVFFNTGGAHRYSHVGLYIGNGEFIHAPGSGKTIKTEKLSTPFYAKNYLGAHTFFTE
                    130       140       150       160       170
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1026>:

```
a206.seq
  1 ATGTTTCCCC CCGACAAAAC CCTTTTCCTC TGTCTCAGCG CACTGCTCCT
 51 CGCCTCATGC GGCACGACCT CCGGCAAACA CCGCCAACCG AAACCCAAAC
101 AGACAGTCCG GCAAATCCAA GCCGTCCGCA TCAGCCACAT CGACCGCACA
151 CAAGGCTCGC AGGAACTCAT GCTCCACAGC CTCGGACTCA TCGGCACGCC
201 CTACAAATGG GGCGGCAGCA GCACCGCAAC CGGCTTCGAT TGCAGCGGCA
251 TGATTCAATT CGTTTACAAA AACGCCCTCA ACGTCAAGCT GCCGCGCACC
301 GCCCGCGACA TGGCGGCGGC AAGCCGCAAA ATCCCCGACA GCCGCCTTAA
351 GGCCGGCGAC CTCGTATTCT TCAACACCGG CGGCGCACAC CGCTACTCAC
401 ACGTCGGACT CTATATCGGC AACGGCGAAT TCATCCATGC CCCCAGCAGC
451 GGCAAAACCA TCAAAACCGA AAAACTCTCC ACACCGTTTT ACGCCAAAAA
501 CTACCTCGGC GCACATACTT TCTTTACAGA ATGA
```

This corresponds to the amino acid sequence <SEQ ID 1027; ORF 206.a>:

```
a206.pep
```

```
  1  MFPPDKTLFL CLSALLLASC GTTSGKHRQP KPKQTVRQIQ AVRISHIDRT
 51  QGSQELMLHS LGLIGTPYKW GGSSTATGFD CSGMIQFVYK NALNVKLPRT
101  ARDMAAASRK IPDSRLKAGD LVFFNTGGAH RYSHVGLYIG NGEFIHAPSS
151  GKTIKTEKLS TPFYAKNYLG AHTFFTE*
```

m206/a206 ORFs 206 and 206.a showed a 99.4% identity in 177 aa overlap

```
                     10        20        30        40        50        60
m206.pep    MFPPDKTLFLCLSALLLASCGTTSGKHRQPKPKQTVRQIQAVRISHIDRTQGSQELMLHS
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a206        MFPPDKTLFLCLSALLLASCGTTSGKHRQPKPKQTVRQIQAVRISHIDRTQGSQELMLHS
                     10        20        30        40        50        60

                     70        80        90       100       110       120
m206.pep    LGLIGTPYKWGGSSTATGFDCSGMIQFVYKNALNVKLPRTARDMAAASRKIPDSRXKAGD
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
a206        LGLIGTPYKWGGSSTATGFDCSGMIQFVYKNALNVKLPRTARDMAAASRKIPDSRLKAGD
                     70        80        90       100       110       120

                    130       140       150       160       170
m206.pep    LVFFNTGGAHRYSHVGLYIGNGEFIHAPSSGKTIKTEKLSTPFYAKNYLGAHTFFTEX
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a206        LVFFNTGGAHRYSHVGLYIGNGEFIHAPSSGKTIKTEKLSTPFYAKNYLGAHTFFTEX
                    130       140       150       160       170
```

287

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1028>:

```
m287.seq
   1  ATGTTTAAAC GCAGCGTAAT CGCAATGGCT TGTATTTTTG CCCTTTCAGC
  51  CTGCGGGGGC GGCGGTGGCG GATCGCCCGA TGTCAAGTCG GCGGACACGC
 101  TGTCAAAACC TGCCGCCCCT GTTGTTTCTG AAAAAGAGAC AGAGGCAAAG
 151  GAAGATGCGC CACAGGCAGG TTCTCAAGGA CAGGGCGCGC CATCCGCACA
 201  AGGCAGTCAA GATATGGCGG CGGTTTCGGA AGAAAATACA GGCAATGGCG
 251  GTGCGGTAAC AGCGGATAAT CCCAAAAATG AAGACGAGGT GGCACAAAAT
 301  GATATGCCGC AAAATGCCGC CGGTACAGAT AGTTCGACAC CGAATCACAC
 351  CCCGGATCCG AATATGCTTG CCGGAAATAT GGAAAATCAA GCAACGGATG
 401  CCGGGGAATC GTCTCAGCCG GCAAACCAAC CGGATATGGC AAATGCGGCG
 451  GACGGAATGC AGGGGGACGA TCCGTCGGCA GGCGGGCAAA ATGCCGGCAA
 501  TACGGCTGCC CAAGGTGCAA ATCAAGCCGG AAACAATCAA GCCGCCGGTT
 551  CTTCAGATCC CATCCCCGCG TCAAACCCTG CACCTGCGAA TGGCGGTAGC
 601  AATTTTGGAA GGGTTGATTT GGCTAATGGC GTTTTGATTG ACGGGCCGTC
 651  GCAAAATATA ACGTTGACCC ACTGTAAAGG CGATTCTTGT AGTGGCAATA
 701  ATTTCTTGGA TGAAGAAGTA CAGCTAAAAT CAGAATTTGA AAAATTAAGT
 751  GATGCAGACA AAATAAGTAA TTACAAGAAA GATGGGAAGA ATGATAAATT
 801  TGTCGGTTTG GTTGCCGATA GTGTGCAGAT GAAGGGAATC AATCAATATA
 851  TTATCTTTTA TAAACCTAAA CCCACTTCAT TTGCGCGATT TAGGCGTTCT
 901  GCACGGTCGA GGCGGTCGCT TCCGGCCGAG ATGCCGCTGA TTCCCGTCAA
 951  TCAGGCGGAT ACGCTGATTG TCGATGGGGA AGCGGTCAGC CTGACGGGGC
1001  ATTCCGGCAA TATCTTCGCG CCCGAAGGGA ATTACCGGTA TCTGACTTAC
1051  GGGGCGGAAA AATTGCCCGG CGGATCGTAT GCCCTTCGTG TTCAAGGCGA
1101  ACCGGCAAAA GGCGAAATGC TTGCGGGCGC GGCCGTGTAC AACGGCGAAG
1151  TACTGCATTT CCATACGGAA AACGGCCGTC CGTACCCGAC CAGGGGCAGG
1201  TTTGCCGCAA AAGTCGATTT CGGCAGCAAA TCTGTGGACG GCATTATCGA
1251  CAGCGGCGAT GATTTGCATA TGGGTACGCA AAAATTCAAA GCCGCCATCG
1301  ATGGAAACGG CTTTAAGGGG ACTTGGACGG AAAATGGCAG CGGGGATGTT
1351  TCCGGAAAGT TTACGGCCCC GGCCGGCGAG GAAGTGGCGG GAAAATACAG
1401  CTATCGCCCG ACAGATGCGG AAAAGGGCGG ATTCGGCGTG TTTGCCGGCA
```

```
1451  AAAAAGAGCA GGATTGA
```

This corresponds to the amino acid sequence <SEQ ID 1029; ORF 287>:

```
m287.pep
   1  MFKRSVIAMA CIFALSACGG GGGGSPDVKS ADTLSKPAAP VVSEKETEAK
  51  EDAPQAGSQG QGAPSAQGSQ DMAAVSEENT GNGGAVTADN PKNEDEVAQN
 101  DMPQNAAGTD SSTPNHTPDP NMLAGNMENQ ATDAGESSQP ANQPDMANAA
 151  DGMQGDDPSA GGQNAGNTAA QGANQAGNNQ AAGSSDPIPA SNPAPANGGS
 201  NFGRVDLANG VLIDGPSQNI TLTHCKGDSC SGNNFLDEEV QLKSEFEKLS
 251  DADKISNYKK DGKNDKFVGL VADSVQMKGI NQYIIFYKPK PTSFARFRRS
 301  ARSRRSLPAE MPLIPVNQAD TLIVDGEAVS LTGHSGNIFA PEGNYRYLTY
 351  GAEKLPGGSY ALRVQGEPAK GEMLAGAAVY NGEVLHFHTE NGRPYPTRGR
 401  FAAKVDFGSK SVDGIIDSGD DLHMGTQKFK AAIDGNGFKG TWTENGSGDV
 451  SGKFYGPAGE EVAGKYSYRP TDAEKGGFGV FAGKKEQD*
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1030>:

```
g287.seq
   1  atgtttaaac gcagtgtgat tgcaatggct tgtatttttc ccctttcagc
  51  ctgtgggggc ggcggtggcg gatcgcccga tgtcaagtcg gcggacacgc
 101  cgtcaaaacc ggccgccccc gttgttgctg aaaatgccgg ggaaggggtg
 151  ctgccgaaag aaaagaaaga tgaggaggca gcgggcggtg cgccgcaagc
 201  cgatacgcag gacgcaaccg ccggagaagg cagccaagat atggcggcag
 251  tttcggcaga aaatacaggc aatggcggtg cggcaacaac ggacaacccc
 301  aaaaatgaag acgcgggggc gcaaaatgat atgccgcaaa atgccgccga
 351  atccgcaaat caaacaggga acaaccaacc cgccggttct tcagattccg
 401  cccccgcgtc aaaccctgcc cctgcgaatg gcggtagcga ttttggaagg
 451  acgaacgtgg gcaattctgt tgtgattgac ggaccgtcgc aaaatataac
 501  gttgacccac tgtaaaggcg attcttgtaa tggtgataat ttattggatg
 551  aagaagcacc gtcaaaatca gaatttgaaa aattaagtga tgaagaaaaa
 601  attaagcgat ataaaaaaga cgagcaacgg gagaattttg tcggtttggt
 651  tgctgacagg gtaaaaaagg atggaactaa caaatatatc atcttctata
 701  cggacaaacc acctactcgt tctgcacggt cgaggaggtc gcttccggcc
 751  gagattccgc tgattcccgt caatcaggcc gatacgctga ttgtggatgg
 801  ggaagcggtc agcctgacgg ggcattccgg caatatcttc gcgcccgaag
 851  ggaattaccg gtatctgact tacggggcgg aaaaattgcc cggcggatcg
 901  tatgccctcc gtgtgcaagg cgaaccggca aaaggcgaaa tgcttgttgg
 951  cacggccgtg tacaacggcg aagtgctgca tttccatatg gaaaacggcc
1001  gtccgtaccc gtccggaggc aggtttgccg caaaagtcga tttcggcagc
1051  aaatctgtgg acggcattat cgacagcggc gatgatttgc atatgggtac
1101  gcaaaaattc aaagccgcca tcgatggaaa cggctttaag ggagcttgga
1151  cggaaaatgg cggcggggat gtttccggaa ggttttacgg cccggccggc
1201  gaggaagtgg cgggaaaata cagctatcgc ccgacagatg ctgaaaaggg
1251  cggattcggc gtgtttgccg gcaaaaaaga tcgggattga
```

This corresponds to the amino acid sequence <SEQ ID 1031; ORF 287.ng>:

```
g287.pep
   1  MFKRSVIAMA CIFPLSACGG GGGGSPDVKS ADTPSKPAAP VVAENAGEGV
  51  LPKEKKDEEA AGGAPQADTQ DATAGEGSQD MAAVSAENTG NGGAATTDNP
 101  KNEDAGAQND MPQNAAESAN QTGNNQPAGS SDSAPASNPA PANGGSDFGR
 151  TNVGNSVVID GPSQNITLTH CKGDSCNGDN LLDEEAPSKS EFEKLSDEEK
 201  IKRYKKDEQR ENFVGLVADR VKKDGTNKYI IFYTDKPPTR SARSRRSLPA
 251  EIPLIPVNQA DTLIVDGEAV SLTGHSGNIF APEGNYRYLT YGAEKLPGGS
 301  YALRVQGEPA KGEMLVGTAV YNGEVLHFHM ENGRPYPSGG RFAAKVDFGS
 351  KSVDGIIDSG DDLHMGTQKF KAAIDGNGFK GTWTENGGGD VSGRFYGPAG
 401  EEVAGKYSYR PTDAEKGGFG VFAGKKDRD*
```

m287/g287 ORFs 287 and 287.ng showed a 70.1% identity in 499 aa overlap

```
                        10        20        30        40                  49
m287.pep    MFKRSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVSE-----------KETEA
            ||||||||||||| ||||||||||||||||||| ||||||||:|           |: ||
g287        MFKRSVIAMACIFPLSACGGGGGGSPDVKSADTPSKPAAPVVAENAGEGVLPKEKKDEEA
                    10        20        30        40        50        60

           50        60        70        80        90       100       109
m287.pep    KEDAPQAGSQGQGAPSAQGSQDMAAVSEENTGNGGAVTADNPKNEDEVAQNDMPQNAAGT
               |||| :|   | :::||||||||| ||||||||:|:|||||||  ||||||||||
g287        AGGAPQADTQD--ATAGEGSQDMAAVSAENTGNGGAATTDNPKNEDAGAQNDMPQNAA--
                    70        80        90       100       110

          110       120       130       140       150       160       169
m287.pep    DSSTPNHTPDPNMLAGNMENQATDAGESSQPANQPDMANAADGMQGDDPSAGGQNAGNTA

g287        ------------------------------------------------------------


          170       180       190       200       210       220       229
m287.pep    AQGANQAGNNQAAGSSDPIPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDS
             ::|||:|||| |||||  ||||||||||||:|||::::|:|:|||||||||||||||||
g287        -ESANQTGNNQPAGSSDSAPASNPAPANGGSDFGRTNVGNSVVIDGPSQNITLTHCKGDS
              120       130       140       150       160       170

          230       240       250       260       270       280       289
m287.pep    CSGNNFLDEEVQLKSEFEKLSDADKISNYKKDGKNDKFVGLVADSVQMKGINQYIIFYKP
            |:|:|:||||:  ||||||||| :||: |||| : ::||||||| |:  | |:|||||
g287        CNGDNLLDEEAPSKSEFEKLSDEEKIKRYKKDEQRENFVGLVADRVKKDGTNKYIIFYTD
              180       190       200       210       220       230

          290       300       310       320       330       340       349
m287.pep    KPTSFARFRRSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLT
            || :     ||||||||||||:||||||||||||||||||||||||||||||||||||||
g287        KPPT-----RSARSRRSLPAEIPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLT
                     240       250       260       270       280       290

          350       360       370       380       390       400       409
m287.pep    YGAEKLPGGSYALRVQGEPAKGEMLAGAAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGS
            |||||||||||||||||||||||||:|:||||||||||| |||||||: |||||||||||
g287        YGAEKLPGGSYALRVQGEPAKGEMLVGTAVYNGEVLHFHMENGRPYPSGGRFAAKVDFGS
                     300       310       320       330       340       350

          410       420       430       440       450       460       469
m287.pep    KSVDGIIDSGDDLHMGTQKFKAAIDGNGFKGTWTENGSGDVSGKFYGPAGEEVAGKYSYR
            |||||||||||||||||||||||||||||||||||||:|||||:||||||||||||||||
g287        KSVDGIIDSGDDLHMGTQKFKAAIDGNGFKGTWTENGGGDVSGRFYGPAGEEVAGKYSYR
                     360       370       380       390       400       410

          470       480       489
m287.pep    PTDAEKGGFGVFAGKKEQDX
            ||||||||||||||||::||
g287        PTDAEKGGFGVFAGKKDRDX
                     420       430
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1032>:

```
a287.seq
      1  ATGTTTAAAC GCAGTGTGAT TGCAATGGCT TGTATTGTTG CCCTTTCAGC
     51  CTGTGGGGGC GGCGGTGGCG GATCGCCCGA TGTTAAGTCG GCGGACACGC
    101  TGTCAAAACC TGCCGCCCCT GTTGTTACTG AAGATGTCGG GGAAGAGGTG
    151  CTGCCGAAAG AAAAGAAAGA TGAGGAGGCG GTGAGTGGTG CGCCGCAAGC
    201  CGATACGCAG GACGCAACCG CCGGAAAAGG CGGTCAAGAT ATGGCGGCAG
    251  TTTCGGCAGA AAATACAGGC AATGGCGGTG CGGCAACAAC GGATAATCCC
```

```
 301  GAAAATAAAG ACGAGGGACC GCAAAATGAT ATGCCGCAAA ATGCCGCCGA
 351  TACAGATAGT TCGACACCGA ATCACACCCC TGCACCGAAT ATGCCAACCA
 401  GAGATATGGG AAACCAAGCA CCGGATGCCG GGGAATCGGC ACAACCGGCA
 451  AACCAACCGG ATATGGCAAA TGCGGCGGAC GGAATGCAGG GGGACGATCC
 501  GTCGGCAGGG GAAAATGCCG GCAATACGGC AGATCAAGCT GCAAATCAAG
 551  CTGAAAACAA TCAAGTCGGC GGCTCTCAAA ATCCTGCCTC TTCAACCAAT
 601  CCTAACGCCA CGAATGGCGG CAGCGATTTT GGAAGGATAA ATGTAGCTAA
 651  TGGCATCAAG CTTGACAGCG GTTCGGAAAA TGTAACGTTG ACACATTGTA
 701  AAGACAAAGT ATGCGATAGA GATTTCTTAG ATGAAGAAGC ACCACCAAAA
 751  TCAGAATTTG AAAAATTAAG TGATGAAGAA AAAATTAATA AATATAAAAA
 801  AGACGAGCAA CGAGAGAATT TTGTCGGTTT GGTTGCTGAC AGGGTAGAAA
 851  AGAATGGAAC TAACAAATAT GTCATCATTT ATAAAGACAA GTCCGCTTCA
 901  TCTTCATCTG CGCGATTCAG GCGTTCTGCA CGGTCGAGGC GGTCGCTTCC
 951  GGCCGAGATG CCGCTGATTC CCGTCAATCA GGCGGATACG CTGATTGTCG
1001  ATGGGGAAGC GGTCAGCCTG ACGGGGCATT CCGGCAATAT CTTCGCGCCC
1051  GAAGGGAATT ACCGGTATCT GACTTACGGG GCGGAAAAAT TGTCCGGCGG
1101  ATCGTATGCC CTCAGTGTGC AAGGCGAACC GGCAAAAGGC GAAATGCTTG
1151  CGGGCACGGC CGTGTACAAC GGCGAAGTGC TGCATTTCCA TATGGAAAAC
1201  GGCCGTCCGT CCCCGTCCGG AGGCAGGTTT GCCGCAAAAG TCGATTTCGG
1251  CAGCAAATCT GTGGACGGCA TTATCGACAG CGGCGATGAT TTGCATATGG
1301  GTACGCAAAA ATTCAAAGCC GTTATCGATG GAAACGGCTT TAAGGGGACT
1351  TGGACGGAAA ATGGCGGCGG GGATGTTTCC GGAAGGTTTT ACGGCCCGGC
1401  CGGCGAAGAA GTGGCGGGAA AATACAGCTA TCGCCCGACA GATGCGGAAA
1451  AGGGCGGATT CGGCGTGTTT GCCGGCAAAA AAGAGCAGGA TTGA
```

This corresponds to the amino acid sequence <SEQ ID 1033; ORF 287.a>:

```
a287.pep
   1  MFKRSVIAMA CIVALSACGG GGGGSPDVKS ADTLSKPAAP VVTEDVGEEV
  51  LPKEKKDEEA VSGAPQADTQ DATAGKGGQD MAAVSAENTG NGGAATTDNP
 101  ENKDEGPQND MPQNAADTDS STPNHTPAPN MPTRDMGNQA PDAGESAQPA
 151  NQPDMANAAD GMQGDDPSAG ENAGNTADQA ANQAENNQVG GSQNPASSTN
 201  PNATNGGSDF GRINVANGIK LDSGSENVTL THCKDKVCDR DFLDEEAPPK
 251  SEFEKLSDEE KINKYKKDEQ RENFVGLVAD RVEKNGTNKY VIIYKDKSAS
 301  SSSARFRRSA RSRRSLPAEM PLIPVNQADT LIVDGEAVSL TGHSGNIFAP
 351  EGNYRYLTYG AEKLSGGSYA LSVQGEPAKG EMLAGTAVYN GEVLHFHMEN
 401  GRPSPSGGRF AAKVDFGSKS VDGIIDSGDD LHMGTQKFKA VIDGNGFKGT
 451  WTENGGGDVS GRFYGPAGEE VAGKYSYRPT DAEKGGFGVF AGKKEQD*
```

```
m287/a287    ORFs 287 and 287.a showed a 77.2% identity in 501 aa overlap

                     10        20        30        40                  49
m287.pep     MFKRSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVSE-----------KETEA
             |||||||||||| |||||||||||||||||||||||||||||:|           |: ||
a287         MFKRSVIAMACIVALSACGGGGGGSPDVKSADTLSKPAAPVVTEDVGEEVLPKEKKDEEA
                     10        20        30        40        50        60

                 50        60        70        80        90       100       109
m287.pep     KEDAPQAGSQGQGAPSAQGSQDMAAVSEENTGNGGAVTADNPKNEDEVAQNDMPQNAAGT
                |||| :|    | :::|:||||||| ||||||:|:|||:|:||  ||||||||| |
a287         VSGAPQADTQ--DATAGKGGQDMAAVSAENTGNGGAATTDNPENKDEGPQNDMPQNAADT
                     70        80        90       100       110

              110       120       130       140       150       160       169
m287.pep     DSSTPNHTPDPNMLAGNMENQATDAGESSQPANQPDMANAADGMQGDDPSAGGQNAGNTA
             |||||||| ||| : :| ||| |||||:||||||||||||||||||||||| :||||||
a287         DSSTPNHTPAPNMPTRDMGNQAPDAGESAQPANQPDMANAADGMQGDDPSAG-ENAGNTA
               120       130       140       150       160       170

              170       180       190       200       210       220       229
m287.pep     AQGANQAGNNQAAGSSDPIPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDS
              |:|||| |||::||::|  ::||  :||||:|||:::|||: :|: |:|:||||||
a287         DQAANQAENNQVGGSQNPASSTNPNATNGGSDFGRINVANGIKLDSGSENVTLTHCKDKV
```

```
                    180       190       200       210       220       230

                  230       240       250       260       270       280       289
m287.pep            CSGNNFLDEEVQLKSEFEKLSDADKISNYKKDGKNDKFVGLVADSVQMKGINQYIIFYKP
                    |:  :|||||:  ||||||||| :||::|||| : ::||||||| |: :| |:|:|:||
a287                CD-RDFLDEEAPPKSEFEKLSDEEKINKYKKDEQRENFVGLVADRVEKNGTNKYVIIYKD
                     240       250       260       270       280       290

                  290       300       310       320       330       340
m287.pep            KP--TSFARFRRSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRY
                    |   :| |||||||||||||||||||||||||||||||||||||||||||||||||||||
a287                KSASSSSARFRRSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRY
                     300       310       320       330       340       350

                    350       360       370       380       390       400
m287.pep            LTYGAEKLPGGSYALRVQGEPAKGEMLAGAAVYNGEVLHFHTENGRPYPTRGRFAAKVDF
                    |||||||| |||||| |||||||||||||:||||||||||| ||||| |: |||||||||
a287                LTYGAEKLSGGSYALSVQGEPAKGEMLAGTAVYNGEVLHFHMENGRPSPSGGRFAAKVDF
                     360       370       380       390       400       410

                    410       420       430       440       450       460
m287.pep            GSKSVDGIIDSGDDLHMGTQKFKAAIDGNGFKGTWTENGSGDVSGKFYGPAGEEVAGKYS
                    ||||||||||||||||||||||||:||||||||||||||:|||||:||||||||||||||
a287                GSKSVDGIIDSGDDLHMGTQKFKAVIDGNGFKGTWTENGGGDVSGRFYGPAGEEVAGKYS
                     420       430       440       450       460       470

                    470       480       489
m287.pep            YRPTDAEKGGFGVFAGKKEQDX
                    ||||||||||||||||||||||
a287                YRPTDAEKGGFGVFAGKKEQDX
                     480       490
```

406

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1034>:

m406.seq

```
     1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
    51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
   101  TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
   151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
   201  CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
   251  TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
   301  GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
   351  TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
   401  CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
   451  ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
   501  CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
   551  GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
   601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
   651  TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
   701  GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
   751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
   801  AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
   851  CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
   901  AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA
   951  AGGACAACCT TGA
```

This corresponds to the amino acid sequence <SEQ ID 1035; ORF 406>:

m406.pep

```
  1  MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251  AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
301  SHEGYGYSDE VVRQHRQGQP *
```

The following partial DNA sequence was identified in *N. gonorrhoeae* <SEQ ID 1036>:

g406.seq

```
  1  ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
101  TCGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201  AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251  TTGATGCACT GATTCGCGGC GAATACATAA ACAGCCCTGC CGTCCGCACC
301  GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351  TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401  CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
451  ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCCGCG
501  CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
551  GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651  TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
701  GAACCAATAA AAAATTGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801  AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTTCTCC GATATCCAAC
851  CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
901  AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC AACATAGACA
951  AGGGCAACCT TGA
```

This corresponds to the amino acid sequence <SEQ ID 1037; ORF 406.ng>:

g406.pep

```
  1  MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSRSSLGLN
151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251  AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
301  SHEGYGYSDE AVRQHRQGQP *
```

ORF 406.ng shows 98.8% identity over a 320 aa overlap with a predicted ORF (ORF406.a) from *N. gonorrhoeae:*

g406/m406

```
                     10        20        30        40        50        60
g406.pep    MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
            |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m406        MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
                     10        20        30        40        50        60

                     70        80        90       100       110       120
g406.pep    KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
m406          KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
                      70        80        90       100       110       120

                     130       140       150       160       170       180
g406.pep      LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
m406          LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
                     130       140       150       160       170       180

                     190       200       210       220       230       240
g406.pep      FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
m406          FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
                     190       200       210       220       230       240

                     250       260       270       280       290       300
g406.pep      IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIQPYGNHTGNSAPSVEADN
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
m406          IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIRPYGNHTGNSAPSVEADN
                     250       260       270       280       290       300

                     310       320
g406.pep      SHEGYGYSDEAVRQHRQGQPX
              ||||||||||:||||||||||
m406          SHEGYGYSDEVVRQHRQGQPX
                     310       320
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1038>:

```
a406.seq
      1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
     51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
    101  TCGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
    151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
    201  AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
    251  TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
    301  GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
    351  TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
    401  CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
    451  ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
    501  CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
    551  GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
    601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
    651  TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
    701  GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
    751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
    801  AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTTCTCC GATATCCAAC
    851  CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
    901  AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA
    951  AGGGCAACCT TGA
```

This corresponds to the amino acid sequence <SEQ ID 1039; ORF 406.a>:

```
a406.pep
      1  MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
     51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
    101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
    151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
    201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
    251  AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
    301  SHEGYGYSDE AVRRHRQGQP *
```

```
m406/a406     ORFs 406 and 406.a showed a 98.8% identity in 320 aa overlap

                        10        20        30        40        50        60
m406.pep      MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a406          MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
                        10        20        30        40        50        60

                        70        80        90       100       110       120
m406.pep      KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a406          KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
                        70        80        90       100       110       120

                       130       140       150       160       170       180
m406.pep      LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a406          LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
                       130       140       150       160       170       180

                       190       200       210       220       230       240
m406.pep      FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
a406          FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
                       190       200       210       220       230       240

                       250       260       270       280       290       300
m406.pep      IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIRPYGNHTGNSAPSVEADN
              ||||||||||||||||||||||||||||||||||||||||:||||| |||||||||||
a406          IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIQPYGNHMGNSAPSVEADN
                       250       260       270       280       290       300

                       310       320
m406.pep      SHEGYGYSDEVVRQHRQGQPX
              ||||||||||:||:|||||||
a406          SHEGYGYSDEAVRRHRQGQPX
                       310       320
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1040>:

```
m726.seq
       1  ATGACCATCT ATTTCAAAAA CGGCTTTTAC GACGACACAT TGGGCGGCAT
      51  CCCCGAAGGC GCGGTTGCCG TCCGCGCCGA AGAATACGCC GCCCTTTTGG
     101  CAGGACAGGC GCAGGGCGGG CAGATTGCCG CAGATTCCGA CGGCCGCCCC
     151  GTTTTAACCC CGCCGCGCCC GTCCGATTAC CACGAATGGG ACGGCAAAAA
     201  ATGGAAAATC AGCAAAGCCG CCGCCGCCGC CCGTTTCGCC AAACAAAAAA
     251  CCGCCTTGGC ATTCCGCCTC GCGGAAAAGG CGGACGAACT CAAAAACAGC
     301  CTCTTGGCGG GCTATCCCCA AGTGGAAATC GACAGCTTTT ACAGGCAGGA
     351  AAAAGAAGCC CTCGCGCGGC AGGCGGACAA CAACGCCCCG ACCCCGATGC
     401  TGGCGCAAAT CGCCGCCGCA AGGGGCGTGG AATTGGACGT TTTGATTGAA
     451  AAAGTTATCG AAAAATCCGC CCGCCTGGCT GTTGCCGCCG GCGCGATTAT
     501  CGGAAAGCGT CAGCAGCTCG AAGACAAATT GAACACCATC GAAACCGCGC
     551  CCGGATTGGA CGCGCTGGAA AAGGAAATCG AAGAATGGAC GCTAAACATC
     601  GGCTGA
```

This corresponds to the amino acid sequence <SEQ ID 1041; ORF 726>:

```
m726.pep
       1  MTIYFKNGFY DDTLGGIPEG AVAVRAEEYA ALLAGQAQGG QIAADSDGRP
      51  VLTPPRPSDY HEWDGKKWKI SKAAAAARFA KQKTALAFRL AEKADELKNS
```

```
101  LLAGYPQVEI DSFYRQEKEA LARQADNNAP TPMLAQIAAA RGVELDVLIE
151  KVIEKSARLA VAAGAIIGKR QQLEDKLNTI ETAPGLDALE KEIEEWTLNI
201  G*
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1042>:

m907-2.seq

```
  1  ATGAGAAAAC CGACCGATAC CCTACCCGTT AATCTGCAAC GCCGCCGCCT
 51  GTTGTGTGCC GCCGGTGCGT TGTTGCTCAG TCCTCTGGCG CACGCCGGCG
101  CGCAACGTGA GGAAACGCTT GCCGACGATG TGGCTTCCGT GATGAGGAGT
151  TCTGTCGGCA GCGTCAATCC GCCGAGGCTG GTGTTTGACA ATCCGAAAGA
201  GGGCGAGCGT TGGTTGTCTG CCATGTCGGC ACGTTTGGCA AGGTTCGTCC
251  CCGAGGAGGA GGAGCGGCGC AGGCTGCTGG TCAATATCCA GTACGAAAGC
301  AGCCGGGCCG GTTTGGATAC GCAGATTGTG TTGGGGCTGA TTGAGGTGGA
351  AAGCGCGTTC CGCCAGTATG CAATCAGCGG TGTCGGCGCG CGCGGCCTGA
401  TGCAGGTTAT GCCGTTTTGG AAAAACTACA TCGGCAAACC GGCGCACAAC
451  CTGTTCGACA TCCGCACCAA CCTGCGTTAC GGCTGTACCA TCCTGCGCCA
501  TTACCGGAAT CTTGAAAAAG GCAACATCGT CCGCGCGCTT GCCCGCTTTA
551  ACGGCAGCTT GGGCAGCAAT AAATATCCGA ACGCCGTTTT GGGCGCGTGG
601  CGCAACCGCT GGCAGTGGCG TTGA
```

This corresponds to the amino acid sequence <SEQ ID 1043; ORF 907-2>:

m907-2.pep

```
  1  MRKPTDTLPV NLQRRRLLCA AGALLLSPLA HAGAQREETL ADDVASVMRS
 51  SVGSVNPPRL VFDNPKEGER WLSAMSARLA RFVPEEEERR RLLVNIQYES
101  SRAGLDTQIV LGLIEVESAF RQYAISGVGA RGLMQVMPFW KNYIGKPAHN
151  LFDIRTNLRY GCTILRHYRN LEKGNIVRAL ARFNGSLGSN KYPNAVLGAW
201  RNRWQWR*
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1044>:

m953.seq

```
  1  ATGAAAAAAA TCATCTTCGC CGCACTCGCA GCCGCCGCCA TCAGTACTGC
 51  CTCCGCCGCC ACCTACAAAG TGGACGAATA TCACGCCAAC GCCCGTTTCG
101  CCATCGACCA TTTCAACACC AGCACCAACG TCGGCGGTTT TTACGGTCTG
151  ACCGGTTCCG TCGAGTTCGA CCAAGCAAAA CGCGACGGTA AAATCGACAT
201  CACCATCCCC ATTGCCAACC TGCAAAGCGG TTCGCAACAC TTTACCGACC
251  ACCTGAAATC AGCCGACATC TTCGATGCCG CCCAATATCC GGACATCCGC
301  TTTGTTTCCA CCAAATTCAA CTTCAACGGC AAAAAACTGG TTTCCGTTGA
351  CGGCAACCTG ACCATGCACG GCAAAACCGC CCCCGTCAAA CTCAAAGCCG
401  AAAAATTCAA CTGCTACCAA AGCCCGATGG AGAAAACCGA AGTTTGTGGC
451  GGCGACTTCA GCACCACCAT CGACCGCACC AAATGGGGCA TGGACTACCT
501  CGTTAACGTT GGTATGACCA AAAGCGTCCG CATCGACATC CAAATCGAGG
551  CAGCCAAACA ATAA
```

This corresponds to the amino acid sequence <SEQ ID 1045; ORF 953>:

m953.pep

```
  1  MKKIIFAALA AAAISTASAA TYKVDEYHAN ARFAIDHFNT STNVGGFYGL
 51  TGSVEFDQAK RDGKIDITIP IANLQSGSQH FTDHLKSADI FDAAQYPDIR
101  FVSTKFNFNG KKLVSVDGNL TMHGKTAPVK LKAEKFNCYQ SPMEKTEVCG
151  GDFSTTIDRT KWGMDYLVNV GMTKSVRIDI QIEAAKQ*
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1046>:

orf1-1.seq

```
   1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
  51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT
 101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
 201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
 351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
 401 ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
 451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAATT
 501 TGTCACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
 551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCGTAT TGGGGCAGGC
 601 AGGCAATATT GGCGATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC
 651 ATATCATATT GCAAGTGCGT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
 701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT
 751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
 801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
 851 ATGGGGTATT GCAAACGGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC
 901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
 951 CCATTCAGTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG
1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
1051 CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACGAA
1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
1251 ATTATATTTC CAAGGAGATT TTACGGTCTC GCCTGAAAAT AACGAAACTT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA
1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAACAGCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGCAGA
1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
1901 CGCAAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
1951 TACAATCATT TAAACGACCA TTGGTCGCAA AAAGAGGGCA TTCCTCGCGG
2001 GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTTGCCAAA
2101 GTGAAAGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTTGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAATTG TGTCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCACGCTCA
2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCCAAAA CGGCAACCTT
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCCTATCG
2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TGCGACAGAT GCGCCGCGCC
2801 GCCGTTCGCG CCGTTCGCGC CGTTCCCTAT TATCCGTTAC ACCGCCAACT
2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
2901 TCAGGGAACA TTCCGCTTTA TGTCGGAACT CTTCGGCTAC CGCAGCGACA
2951 AATTGAAGCT GGCGGAAAGT TCCGAAGGCA CTTACACCTT GGCGGTCAAC
3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAGG
3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTCACC CTGCAAAACG
3101 AACACGTCGA TGCCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC
```

```
3151  GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
3201  CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAAGCC
3251  TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
3301  GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCG GCATTATGCA
3351  GGCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
3401  CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
3451  GCCCGCCGCG CCCGCCGGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC
3501  CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGAGTG
3551  AATTTTCCGC CACGCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601  CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651  GGACACCAAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA
3701  CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751  GGCATCCTGT TTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801  CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA
3851  TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901  AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA
3951  CGGCATTCAG GCACGATACC GCGCCGGTTT CGGCGGATTC GGCATCGAAC
4001  CGCACATCGG CGCAACGCGC TATTTCGTCC AAAAAGCGGA TTACCGCTAC
4051  GAAAACGTCA ATATCGCCAC CCCCGGCCTT GCATTCAACC GCTACCGCGC
4101  GGGCATTAAG GCAGATTATT CATTCAAACC GGCGCAACAC ATTTCCATCA
4151  CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201  ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTCGGCA AAACCCGCAG
4251  TGCGGAATGG GGCGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC
4301  ACGCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351  ATCAAATTAG GCTACCGCTG GTAA
```

This corresponds to the amino acid sequence <SEQ ID 1047; ORF orf1-1>:

orf1-1.pep

```
   1  MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
  51  YQYYRDFAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRNG
 101  VAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
 151  KGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
 201  RQYWRSDEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
 251  KHSPYGFLPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF
 301  QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
 351  LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
 401  GKGELILTSN INQGAGGLYF QGDFTVSPEN NETWQGAGVH ISEDSTVTWK
 451  VNGVANDRLS KIGKGTLHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
 501  FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
 551  DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD
 601  TTKTNGRLNL VYQPAAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651  YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
 701  VKGDWHLSNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
 751  LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
 801  SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
 851  HSALNGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSGTEL
 901  GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
 951  SVESRFNTLT VNGKLNGQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN
1001  NTGNEPASLE QLTVVEGKDN KPLSENLNFT LQNEHVDAGA WRYQLIRKDG
1051  EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101  VAEPARQAGG ENVGIMQAEE EKKRVQADKD TALAKQREAE TRPATTAFPR
1151  ARRARRDLPQ LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201  RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
1251  GILFSHNRTE NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301  SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351  ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR
1401  TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAKG PQLEAQHSAG
1451  IKLGYRW*
```

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1048>:

```
orf46-2.seq
   1  TTGGGCATTT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
  51  CCTGCCGATG CATGCACACG CCTCAGATTT GGCAAACGAT TCTTTTATCC
 101  GGCAGGTTCT CGACCGTCAG CATTTCGAAC CCGACGGGAA ATACCACCTA
 151  TTCGGCAGCA GGGGGGAACT TGCCGAGCGC AGCGGCCATA TCGGATTGGG
 201  AAAAATACAA AGCCATCAGT TGGGCAACCT GATGATTCAA CAGGCGGCCA
 251  TTAAAGGAAA TATCGGCTAC ATTGTCCGCT TTTCCGATCA CGGGCACGAA
 301  GTCCATTCCC CCTTCGACAA CCATGCCTCA CATTCCGATT CTGATGAAGC
 351  CGGTAGTCCC GTTGACGGAT TTAGCCTTTA CCGCATCCAT TGGGACGGAT
 401  ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT
 451  CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT
 501  TGCCCAAAAT ATCCGCCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC
 551  GGCTTGCCGA CCGTTTCCAC AATGCCGGTA GTATGCTGAC GCAAGGAGTA
 601  GGCGACGGAT TCAAACGCGC CACCCGATAC AGCCCCGAGC TGGACAGATC
 651  GGGCAATGCC GCCGAAGCCT TCAACGGCAC TGCAGATATC GTTAAAAACA
 701  TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGCAGGGC
 751  ATAAGCGAAG GCTCAAACAT TGCTGTCATG CACGGCTTGG GTCTGCTTTC
 801  CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
 851  TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
 901  AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT
 951  CCCCATCAAA GGGATTGGAG CTGTTCGGGG AAAATACGGC TTGGGCGGCA
1001  TCACGGCACA TCCTATCAAG CGGTCGCAGA TGGGCGCGAT CGCATTGCCG
1051  AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
1101  ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
1151  GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
1201  AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
1251  TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301  AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
1351  GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
1401  GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAATA AGGAACGGTA
1451  ATATAAACAG TAACTTTAGC CAACATGCTC AACTAGAGAG GGAAATTAAT
1501  AAACTAAAAT CTGCCGATGA AATTAATTTT GCAGATGGAA TGGGAAAATT
1551  TACCGATAGC ATGAATGACA AGGCTTTTAG TAGGCTTGTG AAATCAGTTA
1601  AAGAGAATGG CTTCACAAAT CCAGTTGTGG AGTACGTTGA AATAAATGGA
1651  AAAGCATATA TCGTAAGAGG AAATAATRGG GTTTTTGCTG CAGAATACCT
1701  TGGCAGGATA CATGAATTAA AATTTAAAAA AGTTGACTTT CCTGTTCCTA
1751  ATACTAGTTG GAAAAATCCT ACTGATGTCT TGAATGAATC AGGTAATGTT
1801  AAGAGACCTC GTTATAGGAG TAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 1049; ORF orf46-2>:

```
orf46-2.pep
   1  LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRQ HFEPDGKYHL
  51  FGSRGELAER SGHIGLGKIQ SHQLGNLMIQ QAAIKGNIGY IVRFSDHGHE
 101  VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
 151  PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLADRFH NAGSMLTQGV
 201  GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
 251  ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301  NAAQGIEAVS NIFMAAIPIK GIGAVRGKYG LGGITAHPIK RSQMGAIALP
 351  KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 401  KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
 451  VFDAKPRWEV DRKLNKLTTR EQVEKNVQEI RNGNINSNFS QHAQLEREIN
 501  KLKSADEINF ADGMGKFTDS MNDKAFSRLV KSVKENGFTN PVVEYVEING
 551  KAYIVRGNNR VFAAEYLGRI HELKFKKVDF PVPNTSWKNP TDVLNESGNV
 601  KRPRYRSK*
```

Using the above-described procedures, the following oligonucleotide primers were employed in the polymerase chain reaction (PCR) assay in order to clone the ORFs as indicated:

Oligonucleotides used for PCR

Table 1

ORF	Primer	Sequence	Restriction sites
279	Forward	CGCGGATCCCATATG-TTGCCTGCAATCACGATT <SEQ ID 1050>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAGAAGCGGGCGGCAA <SEQ ID 1051>	XhoI
519	Forward	CGCGGATCCCATATG-TTCAAATCCTTTGTCGTCA <SEQ ID 1052>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGGCGGTTTTGCTGC <SEQ ID 1053>	XhoI
576	Forward	CGCGGATCCCATATG-GCCGCCCCCGCATCT <SEQ ID 1054>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATTTACTTTTTTGATGTCGAC <SEQ ID 1055>	XhoI
919	Forward	CGCGGATCCCATATG-TGCCAAAGCAAGAGCATC <SEQ ID 1056>	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGGCGGTATTCGGG <SEQ ID 1057>	XhoI
121	Forward	CGCGGATCCCATATG-GAAACACAGCTTTACAT <SEQ ID 1058>	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAATAATATCCCGCGCCC <SEQ ID 1059>	XhoI
128	Forward	CGCGGATCCCATATG-ACTGACAACGCACT <SEQ ID 1060>	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACCGCGTTGTCGAAA <SEQ ID 1061>	XhoI
206	Forward	CGCGGATCCCATATG-AAACACCGCCAACCGA <SEQ ID 1062>	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCTGTAAAAAAAGTATGTGC <SEQ ID 1063>	XhoI
287	Forward	CCGGAATTCTAGCTAGC-CTTTCAGCCTGCGGG <SEQ ID 1064>	EcoRI-NheI
	Reverse	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC <SEQ ID 1065>	XhoI
406	Forward	CGCGGATCCCATATG-TGCGGGACACTGACAG <SEQ ID 1066>	BamHI-NdeI

	Reverse	CCCGCTCGAG-AGGTTGTCCTTGTCTATG <SEQ ID 1067>	XhoI

EXAMPLE 2

Expression of ORF 919

The primer described in Table 1 for ORF 919 was used to locate and clone ORF 919. The predicted gene *919* was cloned in pET vector and expressed in *E. coli.* The product of protein expression and purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 919-His fusion protein purification. Mice were immunized with the purified 919-His and sera were used for Western blot (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Symbols: M1, molecular weight marker; PP, purified protein, TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vesicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 919 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 919 are provided in Figure 10. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 919 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 3

Expression of ORF 279

The primer described in Table 1 for ORF 279 was used to locate and clone ORF 279. The predicted gene *279* was cloned in pGex vector and expressed in *E. coli.* The product of protein expression and purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 279-GST purification. Mice were immunized with the purified 279-GST and sera were used for Western blot analysis (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vescicle preparation. Arrows indicate the position of the main recombinant protein product (A) and

the *N. meningitidis* immunoreactive band (B). These experiments confirm that 279 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 279 are provided in Figure 11. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 279 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 4

Expression of ORF 576

The primer described in Table 1 for ORF 576 was used to locate and clone ORF 576. The predicted gene *576* was cloned in pGex vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 576-GST fusion protein purification. Mice were immunized with the purified 576-GST and sera were used for Western blot (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vescicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B).. These experiments confirm that ORF 576 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 576 are provided in Figure 12. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 576 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 5

Expression of ORF 519

The primer described in Table 1 for ORF 519 was used to locate and clone ORF 519. The predicted gene *519* was cloned in pET vector and expressed in *E. coli*. The product of

protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 519-His fusion protein purification. Mice were immunized with the purified 519-His and sera were used for Western blot (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vesicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 519 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 519 are provided in Figure 13. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 519 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 6

Expression of ORF 121

The primer described in Table 1 for ORF 121 was used to locate and clone ORF 121. The predicted gene *121* was cloned in pET vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 121-His fusion protein purification. Mice were immunized with the purified 121-His and sera were used for Western blot analysis (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Results show that 121 is a surface-exposed protein. Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vescicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 121 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 121 are provided in Figure 14. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 121 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 7

Expression of ORF 128

The primer described in Table 1 for ORF 128 was used to locate and clone ORF 128. The predicted gene *128* was cloned in pET vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 128-His purification. Mice were immunized with the purified 128-His and sera were used for Western blot analysis (panel B), FACS analysis (panel C), bactericidal assay (panel D) and ELISA assay (panel E). Results show that 128 is a surface-exposed protein. Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vesicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 128 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 128 are provided in Figure 15. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 128 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 8

Expression of ORF 206

The primer described in Table 1 for ORF 206 was used to locate and clone ORF 206. The predicted gene *206* was cloned in pET vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 206-His purification. Mice were immunized with the purified 206-His and sera were used for Western blot analysis (panel B). It is worthnoting that the immunoreactive band in protein extracts from meningococcus is 38 kDa instead of 17 kDa (panel A). To gain information on the nature of this antibody staining we expressed ORF 206 in *E. coli* without the His-tag and including the predicted leader peptide. Western blot analysis on total protein extracts from *E. coli* expressing this native form of the 206 protein showed a recative band at a position of 38 kDa, as observed in meningococcus. We conclude that the 38 kDa band in panel B) is

specific and that anti-206 antibodies, likely recognize a multimeric protein complex. In panel C is shown the FACS analysis, in panel D the bactericidal assay, and in panel E) the ELISA assay. Results show that 206 is a surface-exposed protein. Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vesicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 206 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 519 are provided in Figure 16. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 206 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 9

Expression of ORF 287

The primer described in Table 1 for ORF 287 was used to locate and clone ORF 287. The predicted gene *287* was cloned in pGex vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 287-GST fusion protein purification. Mice were immunized with the purified 287-GST and sera were used for FACS analysis (panel B), bactericidal assay (panel C), and ELISA assay (panel D). Results show that 287 is a surface-exposed protein. Symbols: M1, molecular weight marker. Arrow indicates the position of the main recombinant protein product (A). These experiments confirm that 287 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 287 are provided in Figure 17. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 287 and the amino acid sequence encoded thereby is provided in Example 1.

EXAMPLE 10

Expression of ORF 406

The primer described in Table 1 for ORF 406 was used to locate and clone ORF 406. The predicted gene *406* was cloned in pET vector and expressed in *E. coli*. The product of protein purification was analyzed by SDS-PAGE. In panel A) is shown the analysis of 406-His fusion protein purification. Mice were immunized with the purified 406-His and sera were used for Western blot analysis (panel B), FACS analysis (panel C), bactericidal assay (panel D), and ELISA assay (panel E). Results show that 406 is a surface-exposed protein. Symbols: M1, molecular weight marker; TP, *N. meningitidis* total protein extract; OMV, *N. meningitidis* outer membrane vescicle preparation. Arrows indicate the position of the main recombinant protein product (A) and the *N. meningitidis* immunoreactive band (B). These experiments confirm that 406 is a surface-exposed protein and that it is a useful immunogen. The hydrophilicity plots, antigenic index, and amphipatic regions of ORF 406 are provided in Figure 18. The AMPHI program is used to predict putative T-cell epitopes (Gao et al 1989, *J. Immunol* 143:3007; Roberts et al. 1996, *AIDS Res Human Retroviruses* 12:593; Quakyi et al. 1992, *Scand J Immunol Suppl* 11:9). The nucleic acid sequence of ORF 406 and the amino acid sequence encoded thereby is provided in Example 1.

The foregoing examples are intended to illustrate but not to limit the invention.

Claims

1. A method for identifying an amino acid sequence, comprising the step of searching for putative open reading frames or protein-coding sequences within one or more of *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

2. A method according to claim 1, comprising the steps of searching a *N. meningitidis* nucleotide sequence for an initiation codon and searching the upstream sequence for an in-frame termination codon.

3. A method for producing a protein, comprising the step of expressing a protein comprising an amino acid sequence identified according to any one of claims 1-2.

4. A method for identifying a protein in *N. mengitidis*, comprising the steps of producing a protein according to claim 3, producing an antibody which binds to the protein, and determining whether the antibody recognises a protein produced by *N. menigitidis*.

5. Nucleic acid comprising an open reading frame or protein-coding sequence identified by a method according to any one of claims 1-2.

6. A protein obtained by the method of claim 3.

7. Nucleic acid comprising one or more of the *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID NO 962 to SEQ ID NO 1044.

8. Nucleic acid comprising a nucleotide sequence having greater than 50% sequence identity to a nucleotide sequence disclosed in the sequence listing SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

9. Nucleic acid comprising a fragment of a nucleotide sequence disclosed in the sequence listing SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

10. Nucleic acid according to claim 9, wherein the fragment is unique to the genome of *N. meningitidis*.

11. Nucleic acid complementary to the nucleic acid of any one of claims 7-10.

12. A protein comprising an amino acid sequence encoded within one or more of the *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

13. A protein comprising an amino acid sequences having greater than 50% sequence identity to an amino acid sequence encoded within one or more of the *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961 and 1068, or even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

14. A protein comprising a fragment of an amino acid sequence selected from the group consisting of one or more odd-numbered SEQ ID NOs 963-1037, amino acid sequences having greater than 50% identity with one or more odd-numbered SEQ ID NOs 963-1045, amino acid sequences encoded within one or more of the *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961 and 1068, and amino acid sequences encoded by one or more even-numbered SEQ ID NOs from SEQ ID 962 to SEQ ID 1044.

15. Nucleic acid encoding a protein according to any one of claims 6-8.

16. A computer, a computer memory, a computer storage medium or a computer database containing the nucleotide sequence of a nucleic acid according to any one of claims 7-11.

17. A computer, a computer memory, a computer storage medium or a computer database containing one or more of the *N. meningitidis* nucleotide sequences SEQ ID NOs 1-961.

18. A polyclonal or monoclonal antibody which binds to a protein according to any one of claims 12-14 or 6.

19. A nucleic acid probe comprising nucleic acid according to any one of claims 5, 7-10, or 15.

20. An amplification primer comprising nucleic acid according to any one of claims 5, 7-10, or 15.

21. A composition comprising (a) nucleic acid according to any one of claims 5, 7-10, or 15; (b) protein according to any one of claims 12-14; and/or (c) an antibody according to claim 18.

22. The use of a composition according to claim 21 as a medicament or as a diagnostic reagent.

23. The use of a composition according to claim 21 in the manufacture of (a) a medicament for treating or preventing infection due to Neisserial bacteria and/or (b) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria.

24. A method of treating a patient, comprising administering to the patient a therapeutically effective amount of a composition according to claim 21.

FIG. 1A

919 (46 kDa)

PURIFICATION

M1 919

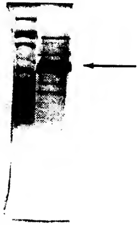


FIG. 1B

919 (46 kDa)

WESTERN BLOT

OMV TP PP

FIG. 1C

919 (46 kDa)

FACS

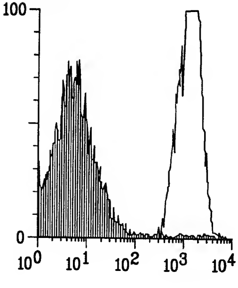


FIG. 1D

919 (46 kDa)

BACTERICIDAL ASSAY

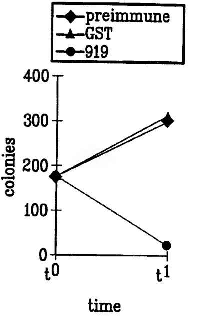


FIG. 1E

919 (46 kDa)

ELISA assay: <u>positive</u>

FIG. 2A

279 (10.5 kDa)

PURIFICATION

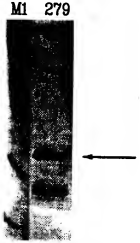


FIG. 2B

279 (10.5 kDa)

WESTERN BLOT

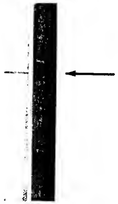


FIG. 2C

279 (10.5 kDa)

FACS

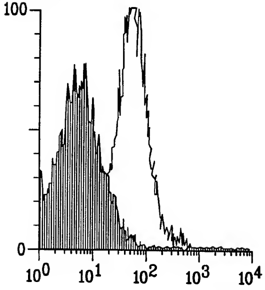


FIG. 2D

279 (10.5 kDa)

BACTERICIDAL ASSAY

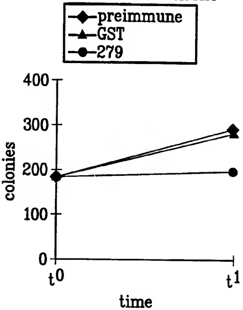


FIG. 2E

279 (10.5 kDa)

ELISA assay: <u>positive</u>

FIG. 3A

576 (27.8 kDa)
PURIFICATION

FIG. 3B

576 (27.8 kDa)
WESTERN BLOT

FIG. 3C

576 (27.8 kDa)
FACS

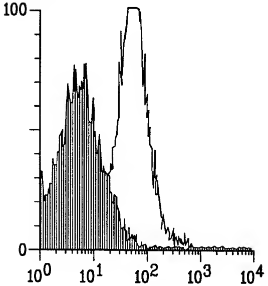


FIG. 3D

576 (27.8 kDa)
BACTERICIDAL ASSAY

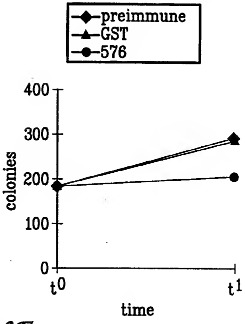


FIG. 3E

576 (27.8 kDa)
ELISA assay: <u>positive</u>

FIG. 4A

519 (33 kDa)
PURIFICATION
M1 519

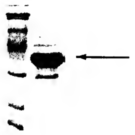


FIG. 4B

519 (33 kDa)
WESTERN BLOT
TP OMV

FIG. 4C

519 (33 kDa)
FACS

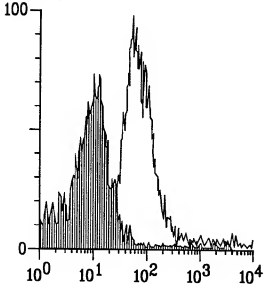


FIG. 4D

519 (33 kDa)
BACTERICIDAL ASSAY

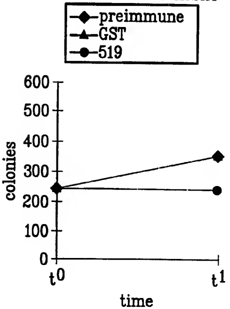


FIG. 4E

519 (33 kDa)
ELISA assay: positive

FIG. 5A

121 (40 kDa)
PURIFICATION

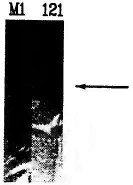


FIG. 5B

121 (40 kDa)
WESTERN BLOT

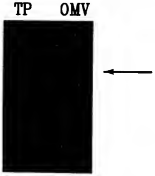


FIG. 5C

121 (40 kDa)
FACS

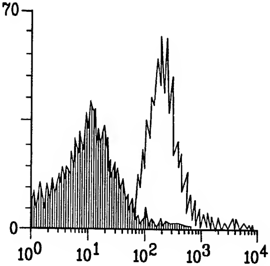


FIG. 5D

121 (40 kDa)
BACTERICIDAL ASSAY

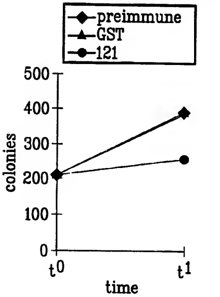


FIG. 5E

121 (40 kDa)

ELISA assay: positive

FIG. 6A

128 (101 kDa)
PURIFICATION
M1 128

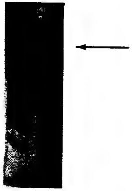


FIG. 6B

128 (101 kDa)
WESTERN BLOT
TP OMV

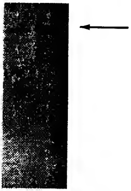


FIG. 6C

128 (101 kDa)
FACS

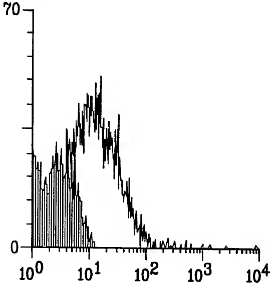


FIG. 6D

128 (101 kDa)
BACTERICIDAL ASSAY

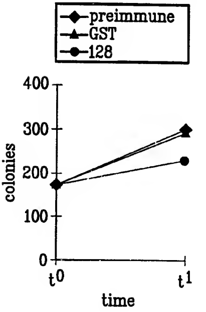


FIG. 6E

128 (101 kDa)
ELISA assay: positive

FIG. 7A

206 (17 kDa)
PURIFICATION

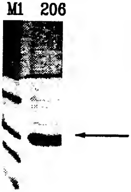


FIG. 7B

206 (17 kDa)

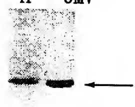


FIG. 7C

206 (17 kDa)
FACS

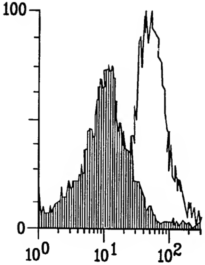


FIG. 7D

206 (17 kDa)
BACTERICIDAL ASSAY

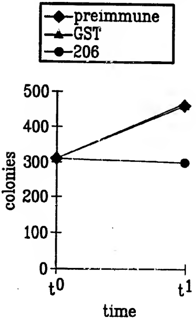


FIG. 7E

206 (17 kDa)

ELISA assay: <u>positive</u>

FIG. 8A

287 (78 kDa)
PURIFICATION

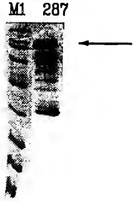


FIG. 8B

287 (78 kDa)
FACS

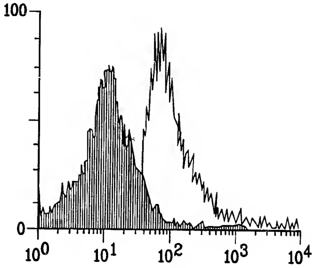


FIG. 8C

287 (78 kDa)
BACTERICIDAL ASSAY

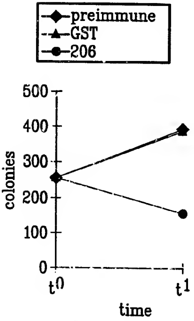


FIG. 8D

287 (78 kDa)
ELISA assay: positive

FIG. 9A

406 (33 kDa)
PURIFICATION
M1 406

FIG. 9B

406 (33 kDa)
WESTERN BLOT
TP OMV

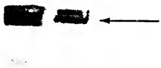


FIG. 9C

406 (33 kDa)
FACS

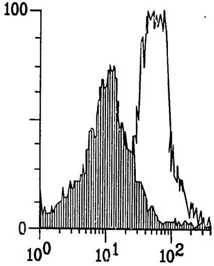


FIG. 9D

406 (33 kDa)
BACTERICIDAL ASSAY

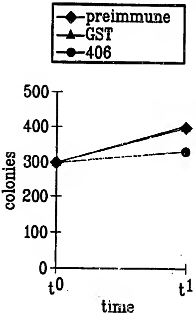


FIG. 9E

406 (33 kDa)

ELISA assay: positive

919
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

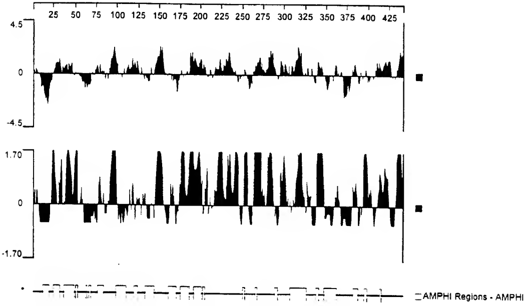


Fig. 10

279

Hydrophilicity Plot, Antigenic Index and AMPHI Regions

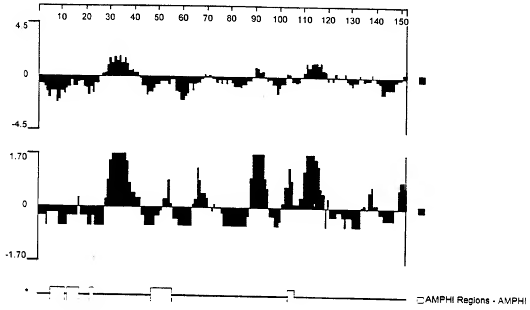


Fig. 11

576-1

Hydrophilicity Plot, Antigenic Index and AMPHI Regions

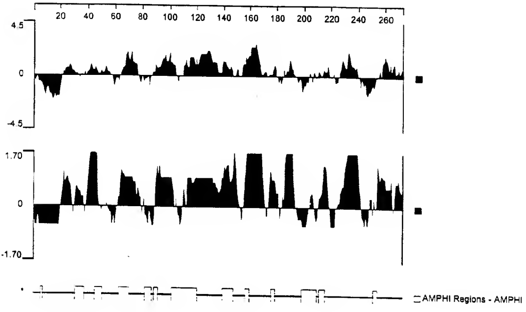


Fig. 12

519-1
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

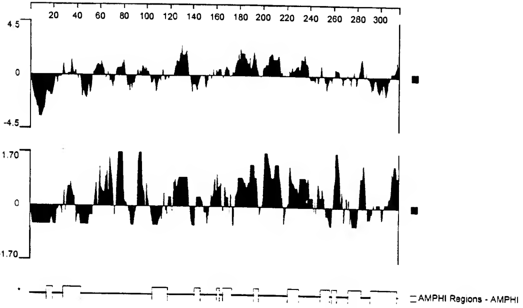


Fig. 13

121-1
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

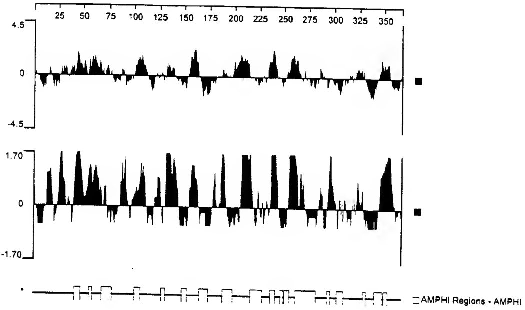


Fig. 14

128-1
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

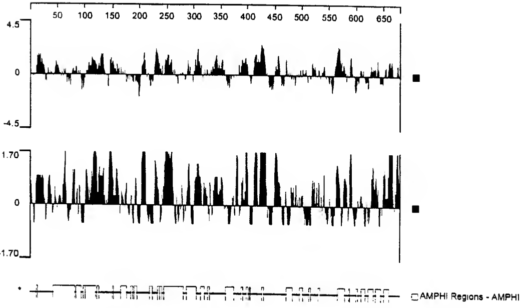


Fig. 15

206
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

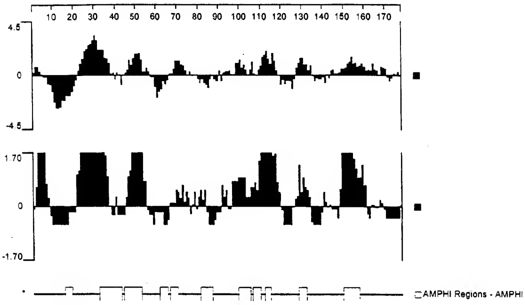


Fig. 16

287

Hydrophilicity Plot, Antigenic Index and AMPHI Regions

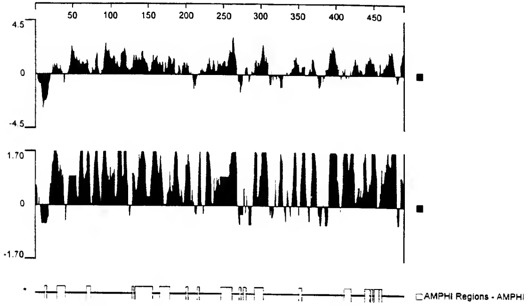


Fig. 17

406
Hydrophilicity Plot, Antigenic Index and AMPHI Regions

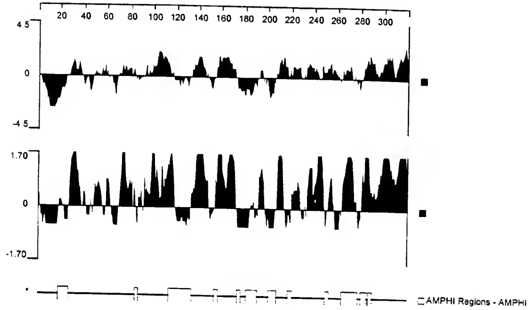


Fig. 18

APPENDIX A

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
1	GNMAA01R	9866	10311
1	GNMAA27F	10765	11284
1	GNMAA27R	11771	12130
1	GNMBA57F	5365	5930
1	GNMBA57R	6594	7118
1	GNMCD17F	9494	10035
1	GNMCD21F	14937	15512
1	GNMCD21R	16217	16700
1	GNMCD26F	27033	27561
1	GNMCD26R	25650	26101
1	GNMCD28F	27012	27561
1	GNMCD58F	27525	28047
1	GNMCD58R	26208	26582
1	GNMCF39F	25928	26411
1	GNMCF39R	24501	25188
1	GNMCK12F	18475	18966
1	GNMCK12R	16734	17175
1	GNMCL43F	31264	31793
1	GNMCL43R	32603	33038
1	GNMCL77F	7112	7681
1	GNMCL77R	8587	9143
1	GNMCO24R	8321	8920
1	GNMCP77F	24906	25412
1	GNMCP77R	26565	27107
1	GNMCQ74F	14937	15617
1	GNMCQ74R	13764	14477
1	GNMCS43F	3607	4278
1	GNMCS56F	21955	22578
1	GNMCS57F	7909	8608
1	GNMCV14F	5771	6272
1	GNMCV15R	7143	7800
1	GNMCV64F	23017	23484
1	GNMCV64R	21277	22018
1	GNMCV74F	16990	17305
1	GNMCV74R	18058	18796
1	GNMCV83F	4008	4503
1	GNMCV83R	2768	3286
1	GNMCY30F	7157	7897
1	GNMCY30R	8378	8912
1	GNMCZ78F	14192	14686
1	GNMCZ78R	15697	16234
1	GNMCZ93F	31337	31862
1	GNMCZ93R	30119	30639
2	GNMAA02F	27133	27648

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
2	GNMAA02R	26120	26546
2	GNMAA38F	16163	16379
2	GNMAA38R	14815	15335
2	GNMAA46F	2337	2704
2	GNMAA46R	3242	3746
2	GNMBA17F	15637	15798
2	GNMCD47F	11113	11453
2	GNMCD78F	13704	14196
2	GNMCD78R	15013	15380
2	GNMCK27F	4941	5490
2	GNMCK27R	3670	4086
2	GNMCL17F	23033	23527
2	GNMCL17R	21424	21995
2	GNMCL82F	24805	25200
2	GNMCL82R	26093	26659
2	GNMCN19F	5929	6601
2	GNMCP32F	18556	19103
2	GNMCP32R	19956	20403
2	GNMCQ84F	16351	17040
2	GNMCQ92F	3243	3692
2	GNMCQ92R	2022	2644
2	GNMCS51F	6645	7300
2	GNMCV24F	28139	28637
2	GNMCV25R	26839	27453
2	GNMCV77F	5149	5575
2	GNMCV77R	6008	6841
2	GNMCY52F	21892	22580
2	GNMCY52R	23157	23662
2	GNMCY74F	21900	22552
2	GNMCY74R	23519	24073
2	GNMCZ69F	1489	1999
2	GNMCZ70F	1489	1985
2	GNMCZ70R	2707	3232
3	GNMAA03F	16946	17459
3	GNMAA03R	18236	18447
3	GNMAA15F	3641	4156
3	GNMAA15R	4704	5176
3	GNMCA12F	8812	9427
3	GNMCB27F	19908	20403
3	GNMCB27R	21309	21630
3	GNMCB59F	22046	22554
3	GNMCB59R	20650	21230
3	GNMCD50F	8711	9229
3	GNMCF53F	15376	15861
3	GNMCF53R	16619	17312
3	GNMCF86F	22322	22760

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
3	GNMCL55F	12659	13194
3	GNMCL55R	13854	14380
3	GNMCM46R	11972	12662
3	GNMCM63F	7397	8071
3	GNMCM63R	8734	9381
3	GNMCP05F	2224	2964
3	GNMCV27F	10472	10969
3	GNMCV28R	11455	12172
4	GNMAA04R	21367	21727
4	GNMAA66F	9998	10514
4	GNMAA66R	9150	9669
4	GNMAA70F	19444	19961
4	GNMAA70R	20446	20841
4	GNMAB18F	34311	34576
4	GNMAB18R	32690	33102
4	GNMBA24F	21408	21950
4	GNMCA71F	35444	36106
4	GNMCA85F	14906	15535
4	GNMCB46F	27141	27652
4	GNMCB46R	28558	29138
4	GNMCD85F	25929	26447
4	GNMCF35F	37587	38065
4	GNMCF35R	36661	37327
4	GNMCK26F	23722	24268
4	GNMCK26R	25176	25751
4	GNMCK39F	26270	26836
4	GNMCK39R	27576	27934
4	GNMCK64F	37686	38053
4	GNMCK64R	36356	36915
4	GNMCL60F	2659	3206
4	GNMCL60R	4028	4520
4	GNMCM12F	21992	22465
4	GNMCM12R	23335	23919
4	GNMCM80F	15507	16171
4	GNMCM80R	16264	16990
4	GNMCN08R	33415	33739
4	GNMCO47F	23101	23700
4	GNMCO47R	24872	25344
4	GNMCP24F	34864	35552
4	GNMCP24R	33620	34225
4	GNMCP44F	24613	24976
4	GNMCP44R	25712	26279
4	GNMCQ80F	35274	35964
4	GNMCQ80R	34053	34632
4	GNMCS02F	37528	38035
4	GNMCV40F	33203	33632

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
4	GNMCX19F	37333	38076
4	GNMCX19R	36229	36871
4	GNMCX25F	28667	29362
4	GNMCX25R	27755	28398
4	GNMCX31F	1336	2085
4	GNMCX31R	1	640
4	GNMCX38F	15063	15774
4	GNMCX38R	14158	14836
4	GNMCY53F	8159	8846
4	GNMCY53R	6905	7405
4	GNMCZ25F	42411	42912
4	GNMCZ25R	40673	41229
4	GNMCZ27F	4786	5245
4	GNMCZ27R	3484	4030
5	GNMAA05F	5819	6334
5	GNMAA05R	6898	7190
5	GNMAA09F	15867	16369
5	GNMAA09R	15935	16368
5	GNMAA50R	17996	18383
5	GNMAA51F	44043	44409
5	GNMAA51R	43157	43679
5	GNMCA06F	43254	43764
5	GNMCA72F	7437	8102
5	GNMCA87F	36458	36899
5	GNMCB41F	44654	45224
5	GNMCB41R	45601	46039
5	GNMCD77F	46927	47437
5	GNMCD77R	48378	48761
5	GNMCF13F	18408	18911
5	GNMCF13R	16858	17553
5	GNMCF26F	44946	45450
5	GNMCF26R	46355	47018
5	GNMCF51F	31870	32355
5	GNMCK15F	34028	34591
5	GNMCK15R	33072	33560
5	GNMCK52F	13042	13587
5	GNMCK52R	11706	12267
5	GNMCK67F	16111	16399
5	GNMCK67R	14116	14459
5	GNMCL36F	26130	26644
5	GNMCL36R	24478	25038
5	GNMCL57F	46883	47459
5	GNMCL57R	48232	48759
5	GNMCL93F	6901	7404
5	GNMCL93R	5298	5897
5	GNMCN22F	4118	4792

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
5	GNMCN22R	5337	5969
5	GNMCN58F	17211	17798
5	GNMCN58R	15825	16436
5	GNMCN85F	38026	38698
5	GNMCN85R	39079	39669
5	GNMCP14F	47197	47893
5	GNMCP14R	47924	48597
5	GNMCP42F	23201	23701
5	GNMCP42R	24295	24875
5	GNMCP60F	31050	31537
5	GNMCP60R	29886	30442
5	GNMCQ39R	321	1003
5	GNMCS18F	39300	39713
5	GNMCS74F	41338	41970
5	GNMCS84F	47085	47801
5	GNMCS85R	48062	48687
5	GNMCV51F	33257	33720
5	GNMCV53F	35594	36106
5	GNMCV53R	36624	37232
5	GNMCV80F	3433	3924
5	GNMCV80R	2239	2949
5	GNMCX14F	15425	16088
5	GNMCX14R	14412	15041
5	GNMCY05F	26090	26786
5	GNMCY05R	25093	25665
5	GNMCY24F	45941	46684
5	GNMCY24R	47197	47748
5	GNMCY75F	9003	9618
5	GNMCY75R	9968	10503
5	GNMCZ74F	32693	33186
5	GNMCZ74R	31650	32179
6	GNMAA06F	43077	43280
6	GNMAA33F	21695	22061
6	GNMAA33R	22761	23120
6	GNMAA39F	11023	11390
6	GNMAA39R	12412	12870
6	GNMAB43F	13579	14098
6	GNMAB56F	20656	21079
6	GNMCA67F	37544	38219
6	GNMCB01F	34331	34902
6	GNMCB01R	35502	36050
6	GNMCD62F	6122	6648
6	GNMCD62R	4831	5183
6	GNMCD93F	1679	2157
6	GNMCD93R	3169	3495
6	GNMCK06F	20928	21478

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
6	GNMCK06R	19697	20289
6	GNMCL39F	24705	25251
6	GNMCL39R	23194	23548
6	GNMCM21F	32432	33056
6	GNMCM21R	33649	34334
6	GNMCN70R	14256	14926
6	GNMCO52F	13197	13922
6	GNMCO85F	26216	26827
6	GNMCO85R	25022	25686
6	GNMCS27F	16689	17300
6	GNMCS61F	3508	4184
6	GNMCS77F	40570	41276
6	GNMCS83F	32447	33093
6	GNMCS84R	30598	31235
6	GNMCV08F	42819	43260
6	GNMCV09R	44363	44932
6	GNMCV75F	14981	15479
6	GNMCX36F	38996	39738
6	GNMCX36R	39855	40528
6	GNMCX59F	39178	39574
6	GNMCX59R	40477	41178
6	GNMCY92F	24695	25185
6	GNMCZ42F	15656	16179
6	GNMCZ42R	17126	17641
6	GNMCZ59F	38912	39364
6	GNMCZ59R	37528	38062
7	GNMAA07F	8291	8808
7	GNMAA07R	9371	9793
7	GNMAA10F	39307	39822
7	GNMAA10R	37810	38060
7	GNMAA76F	289	796
7	GNMAA76R	1117	1517
7	GNMAB01F	33973	34541
7	GNMAB01R	34969	35306
7	GNMAB04F	53611	54157
7	GNMAB04R	52653	53059
7	GNMAB52F	37174	37740
7	GNMAB55F	52123	52618
7	GNMBA81F	28757	29327
7	GNMBA81R	27546	28097
7	GNMBB21F	40393	40959
7	GNMBB21R	39008	39449
7	GNMCA75F	31357	32032
7	GNMCB25F	33514	34085
7	GNMCB25R	34748	35431
7	GNMCB48F	14504	15191

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
7	GNMCB56F	36436	37114
7	GNMCB56R	35390	36079
7	GNMCB67F	42108	42771
7	GNMCB67R	41133	41740
7	GNMCB69F	27142	27807
7	GNMCB69R	25881	26530
7	GNMCD33R	50431	50757
7	GNMCD51F	6134	6629
7	GNMCF11F	35219	35727
7	GNMCF11R	36756	37229
7	GNMCF37F	51876	52358
7	GNMCF37R	49997	50607
7	GNMCF45F	40695	41177
7	GNMCF45R	41795	42403
7	GNMCF58F	6844	7311
7	GNMCF58R	5528	6208
7	GNMCF89F	52016	52469
7	GNMCF89R	53363	54002
7	GNMCH63F	39350	39770
7	GNMCH80F	20170	20607
7	GNMCK02F	43141	43483
7	GNMCK02R	41418	41852
7	GNMCK03F	41843	42407
7	GNMCK03R	40397	40952
7	GNMCK75F	29011	29346
7	GNMCK75R	27279	27840
7	GNMCL37F	37566	38097
7	GNMCL37R	38870	39442
7	GNMCL38F	38465	38990
7	GNMCL38R	37261	37843
7	GNMCL50F	52471	53006
7	GNMCL50R	51307	51879
7	GNMCM16R	43200	43943
7	GNMCM28F	31079	31677
7	GNMCM28R	29986	30699
7	GNMCM75F	29426	30002
7	GNMCM75R	28230	28947
7	GNMCN07R	31678	32296
7	GNMCN08F	30220	30908
7	GNMCN66F	49682	50383
7	GNMCN68R	48507	48702
7	GNMCP52F	53906	54238
7	GNMCP75F	3335	3631
7	GNMCP75R	2430	2916
7	GNMCP87F	19818	20336
7	GNMCP87R	21539	21853

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
7	GNMCQ05F	16992	17629
7	GNMCQ05R	15900	16596
7	GNMCQ06F	8173	8758
7	GNMCQ06R	6774	7461
7	GNMCQ11F	35268	35953
7	GNMCQ11R	36305	36981
7	GNMCQ13F	28320	29037
7	GNMCQ13R	29418	30079
7	GNMCQ24F	40176	40783
7	GNMCQ24R	40841	41510
7	GNMCQ37R	20188	20919
7	GNMCQ55F	40743	41309
7	GNMCQ55R	41980	42698
7	GNMCS30F	49344	49993
7	GNMCS53F	16879	17595
7	GNMCS95F	29469	29622
7	GNMCV01R	30937	31651
7	GNMCV17F	24334	24812
7	GNMCV18R	25368	26100
7	GNMCV28F	26427	26916
7	GNMCV29R	24847	25211
7	GNMCV69F	16647	17098
7	GNMCV91F	10009	10521
7	GNMCV91R	8630	9420
7	GNMCX23F	36634	37387
7	GNMCX23R	38318	38893
7	GNMCX24R	33857	34497
7	GNMCX67F	44537	45096
7	GNMCX67R	45763	46455
7	GNMCX77F	3423	4090
7	GNMCY56F	44117	44788
7	GNMCY56R	45883	46440
7	GNMCY79F	37394	38041
7	GNMCY79R	38954	39287
7	GNMCY84F	7387	8023
7	GNMCY84R	8749	9223
7	GNMCZ21F	28454	28986
7	GNMCZ21R	29774	30347
8	GNMAA08F	3883	4232
8	GNMAA08R	4930	5373
8	GNMAA17F	20102	20622
8	GNMAA17R	19135	19510
8	GNMAA18F	18255	18770
8	GNMAA69F	3985	4501
8	GNMAA69R	2840	3310
8	GNMBA02R	18827	19205

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
8	GNMBA38R	20196	20729
8	GNMBB17F	16245	16809
8	GNMBB17R	14789	15278
8	GNMCD01F	1726	2071
8	GNMCD01R	3032	3560
8	GNMCD57F	15533	16080
8	GNMCD57R	14017	14387
8	GNMCH21F	7735	8074
8	GNMCH58F	20193	20483
8	GNMCK17F	12025	12589
8	GNMCK17R	13519	14068
8	GNMCN37F	11716	12367
8	GNMCN37R	10459	10898
8	GNMCQ71F	15717	16394
8	GNMCQ71R	17082	17799
8	GNMCV56F	2818	3221
8	GNMCV56R	4184	4873
8	GNMCW18F	11443	12002
8	GNMCW19F	12243	12874
8	GNMCX44F	13230	13907
8	GNMCX44R	12093	12776
8	GNMCX81F	6904	7509
8	GNMCX81R	8613	9312
9	GNMAA11R	3820	4070
9	GNMCF10F	4237	4718
9	GNMCF10R	5381	6021
9	GNMCF16F	6231	6723
9	GNMCF16R	4976	5578
9	GNMCH10F	8003	8324
9	GNMCH10R	6412	6686
9	GNMCS36F	8057	8725
9	GNMCX89R	7787	8447
10	GNMAA12F	700	1214
11	GNMAA13F	48121	48639
11	GNMAA13R	49787	50045
11	GNMAA73F	9309	9827
11	GNMAA73R	10319	10725
11	GNMAA95F	5068	5583
11	GNMAA95R	4340	4731
11	GNMAB70F	44475	44906
11	GNMAB70R	45692	46213
11	GNMAB84F	34949	35517
11	GNMAB84R	35628	36115
11	GNMBA30F	35071	35637
11	GNMBA30R	34080	34618
11	GNMBA65F	46358	46779

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
11	GNMBA65R	48334	48629
11	GNMBA96F	25616	26168
11	GNMBA96R	27180	27576
11	GNMCA79F	12432	13093
11	GNMCA81F	64372	65033
11	GNMCB75F	12474	13003
11	GNMCB75R	11368	11898
11	GNMCB79F	12463	12998
11	GNMCB79R	11374	11879
11	GNMCB80F	12394	13044
11	GNMCB80R	11355	11761
11	GNMCB88F	26453	27107
11	GNMCB88R	25225	25878
11	GNMCD37R	1837	2210
11	GNMCD48F	36014	36541
11	GNMCD48R	37485	37833
11	GNMCD61F	33776	34331
11	GNMCD61R	32513	32886
11	GNMCF05F	61923	62430
11	GNMCF05R	63324	63994
11	GNMCF20F	64093	64548
11	GNMCF20R	62670	63312
11	GNMCF27F	7865	8322
11	GNMCF27R	6252	6941
11	GNMCF31F	2643	3144
11	GNMCF31R	3621	4255
11	GNMCF32F	34812	35310
11	GNMCF32R	33489	34167
11	GNMCF44F	7905	8323
11	GNMCF44R	6275	6806
11	GNMCF54F	4208	4682
11	GNMCF54R	5789	6419
11	GNMCH29F	4781	5137
11	GNMCH75F	60773	61203
11	GNMCH75R	62111	62403
11	GNMCK80F	40661	41202
11	GNMCK80R	39298	39847
11	GNMCL01F	59052	59569
11	GNMCL01R	57689	58283
11	GNMCL62F	36623	37174
11	GNMCL62R	38138	38721
11	GNMCL65F	11758	12282
11	GNMCL65R	13221	13807
11	GNMCM44R	3393	4077
11	GNMCM85R	60497	61118
11	GNMCN29F	75370	76048

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
11	GNMCN29R	76487	77001
11	GNMCN90F	53115	53836
11	GNMCN90R	51986	52525
11	GNMCP26F	38602	39106
11	GNMCP26R	37257	37549
11	GNMCQ58F	61396	62055
11	GNMCQ58R	62637	63355
11	GNMCS12F	7065	7598
11	GNMCV05F	4623	5085
11	GNMCV06R	3299	4083
11	GNMCV16F	51884	52341
11	GNMCV17R	53784	54354
11	GNMCV88F	70556	71043
11	GNMCV88R	69005	69740
11	GNMCW41F	39495	40133
11	GNMCX04F	26396	27141
11	GNMCX04R	25242	25882
11	GNMCX65F	43846	44360
11	GNMCX65R	45795	46258
11	GNMCY01F	42714	43318
11	GNMCY03F	16064	16747
11	GNMCY03R	17171	17665
11	GNMCY76F	36967	37624
11	GNMCY76R	38440	38999
11	GNMCZ26F	45695	46211
11	GNMCZ26R	46903	47445
11	GNMCZ30F	53419	53933
11	GNMCZ30R	54651	55202
11	GNMCZ86R	43568	43996
12	GNMAA14F	51035	51374
12	GNMAA62F	22307	22668
12	GNMAA62R	21211	21585
12	GNMAA84F	4132	4648
12	GNMAA84R	3028	3497
12	GNMAB19F	53197	53641
12	GNMAB19R	51715	51941
12	GNMAB34F	59820	60248
12	GNMAB75F	8230	8726
12	GNMAB75R	6772	7086
12	GNMBA16F	61880	62448
12	GNMBA16R	63397	63930
12	GNMBA55F	54894	55463
12	GNMBA55R	53249	53699
12	GNMBB07F	45401	45967
12	GNMBB07R	46474	46846
12	GNMBB23F	23330	23896

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
12	GNMBB23R	21762	22258
12	GNMBB28F	17524	18093
12	GNMBB28R	19255	19581
12	GNMCA08F	80267	80572
12	GNMCA26F	95492	95876
12	GNMCB71F	3761	4447
12	GNMCB71R	2760	3305
12	GNMCD40F	25822	26340
12	GNMCD40R	27392	27712
12	GNMCF14F	254	698
12	GNMCF23F	25032	25512
12	GNMCF23R	26296	26954
12	GNMCF59F	543	781
12	GNMCF59R	1909	2359
12	GNMCF75F	38537	38993
12	GNMCH09F	70027	70360
12	GNMCH09R	68764	69057
12	GNMCK63F	82010	82461
12	GNMCK63R	83284	83844
12	GNMCL27F	36594	37139
12	GNMCL27R	38339	38900
12	GNMCL83F	24969	25304
12	GNMCL83R	26594	27175
12	GNMCM24F	58035	58620
12	GNMCM24R	56788	57519
12	GNMCM26R	43862	44449
12	GNMCM33F	59354	60069
12	GNMCM33R	58194	58939
12	GNMCN23F	31658	32330
12	GNMCN23R	29999	30623
12	GNMCP07F	62762	63498
12	GNMCP07R	61716	62463
12	GNMCQ25F	29033	29713
12	GNMCQ25R	27952	28642
12	GNMCQ31F	33826	34489
12	GNMCQ31R	32628	33318
12	GNMCQ35F	99046	99645
12	GNMCQ35R	100151	100867
12	GNMCS06F	35210	35790
12	GNMCS07F	38327	38874
12	GNMCS37F	93209	93927
12	GNMCS45F	52207	52867
12	GNMCS59F	49955	50647
12	GNMCS63F	13556	14245
12	GNMCS75F	95191	95899
12	GNMCS94F	39007	39638

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
12	GNMCV02F	96642	97004
12	GNMCV03R	95290	96043
12	GNMCV19F	13169	13632
12	GNMCV20R	11334	12063
12	GNMCV67F	12472	12929
12	GNMCV67R	11158	11877
12	GNMCV95F	48011	48518
12	GNMCV95R	48642	49450
12	GNMCX03F	64105	64613
12	GNMCX03R	65502	66139
12	GNMCX62F	91416	91831
12	GNMCX68R	55716	56405
12	GNMCX82F	55372	56082
12	GNMCX82R	54147	54839
12	GNMCX90F	81959	82454
12	GNMCX90R	83099	83791
12	GNMCX91F	82087	82392
12	GNMCY47F	80254	80920
12	GNMCY47R	78886	79381
12	GNMCY81F	17736	18413
12	GNMCY81R	19180	19621
12	GNMCZ02F	24891	25412
12	GNMCZ02R	26406	26946
12	GNMCZ10F	34243	34706
12	GNMCZ10R	35555	36086
12	GNMCZ54F	59674	60174
12	GNMCZ54R	58180	58651
12	GNMCZ65F	70323	70828
12	GNMCZ65R	71871	72382
13	GNMAA19F	12931	13449
13	GNMAA19R	11822	12291
13	GNMAA55R	4581	5101
13	GNMAA63F	36862	37225
13	GNMAA63R	35706	36096
13	GNMAA77F	20561	20750
13	GNMAB20F	14416	14852
13	GNMBA41R	21126	21626
13	GNMCB15F	3423	3980
13	GNMCB15R	4343	4984
13	GNMCB38F	22717	23346
13	GNMCB38R	21451	22022
13	GNMCB57F	11695	12343
13	GNMCD23F	33967	34506
13	GNMCD23R	32498	32984
13	GNMCD27F	25756	26330
13	GNMCD27R	24266	24695

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
13	GNMCD30F	25823	26369
13	GNMCD30R	24703	25016
13	GNMCD91F	36457	36958
13	GNMCF77F	11321	11777
13	GNMCF77R	9878	10580
13	GNMCH04F	9222	9510
13	GNMCK07F	20658	21162
13	GNMCK07R	21983	22516
13	GNMCK24F	11029	11566
13	GNMCK24R	12531	12904
13	GNMCL26F	33412	33883
13	GNMCL26R	32004	32585
13	GNMCL42F	25017	25487
13	GNMCL42R	26410	26988
13	GNMCM18F	9081	9580
13	GNMCM18R	7774	8463
13	GNMCM79F	28296	28959
13	GNMCM79R	29623	30321
13	GNMCN57F	43959	44583
13	GNMCN57R	42560	43109
13	GNMCO81F	36053	36717
13	GNMCO81R	34853	35488
13	GNMCP18F	20932	21612
13	GNMCP18R	19724	20394
13	GNMCS73F	26639	27284
13	GNMCS76R	25539	26264
13	GNMCV09F	46801	47242
13	GNMCV10R	45342	46019
13	GNMCV48F	40436	40867
13	GNMCV81F	21352	21853
13	GNMCW37F	45183	45820
13	GNMCX11F	1628	2393
13	GNMCX11R	2983	3629
13	GNMCX76F	41236	41920
13	GNMCX76R	42308	42978
13	GNMCY20F	20524	21188
13	GNMCY20R	19350	19922
13	GNMCY46F	15097	15751
13	GNMCY46R	16501	17054
13	GNMCY87F	21699	22313
13	GNMCY87R	20274	20660
13	GNMCZ29F	46571	47106
14	GNMAA20F	2883	3399
15	GNMAA21F	12719	13236
15	GNMAA21R	11967	12439
15	GNMAA83F	2799	3318

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
15	GNMAA83R	3978	4448
15	GNMBA09F	4054	4621
15	GNMCB52F	15275	16007
15	GNMCB52R	16498	16827
15	GNMCB77F	18627	19229
15	GNMCB77R	20264	20766
15	GNMCB83F	18623	19271
15	GNMCB83R	20266	20777
15	GNMCL14F	3072	3593
15	GNMCL14R	1651	2228
15	GNMCL87R	9692	10245
15	GNMCN52F	5357	5991
15	GNMCN52R	6753	7339
15	GNMCP45F	11548	12079
15	GNMCP45R	13429	13801
15	GNMCQ09F	19788	20364
15	GNMCQ09R	18441	19134
15	GNMCQ40F	20922	21572
15	GNMCQ40R	22245	22939
15	GNMCV26F	13405	13894
15	GNMCV27R	12194	12828
15	GNMCW08F	23327	23910
15	GNMCX17F	4323	5048
15	GNMCX17R	3040	3690
16	GNMAA22F	54115	54632
16	GNMAA22R	55087	55557
16	GNMAA40R	44790	45219
16	GNMAA72F	58127	58639
16	GNMAA72R	57179	57650
16	GNMAB05F	47515	48081
16	GNMAB05R	46674	47004
16	GNMAB06F	65453	66020
16	GNMAB06R	66416	66833
16	GNMAB07F	65453	65772
16	GNMAB28F	70440	71008
16	GNMAB28R	71467	71806
16	GNMAB41F	21694	22260
16	GNMAB54F	45585	46150
16	GNMAB65F	18770	19084
16	GNMBA69F	9418	9986
16	GNMBA69R	8303	8848
16	GNMBA76F	39980	40549
16	GNMBA76R	41451	41944
16	GNMBA79R	1185	1359
16	GNMCA89F	63127	63781
16	GNMCB30F	5241	5748

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
16	GNMCB32R	3919	4495
16	GNMCD69F	20174	20609
16	GNMCD69R	21508	21899
16	GNMCD74F	20264	20751
16	GNMCF08F	25798	26287
16	GNMCF08R	24361	25036
16	GNMCF36R	42733	43371
16	GNMCF46R	4203	4663
16	GNMCF48F	40973	41398
16	GNMCF48R	39629	40232
16	GNMCF73F	27684	28143
16	GNMCF73R	26442	27127
16	GNMCF81F	67923	68332
16	GNMCH17F	68971	69291
16	GNMCH34R	22199	22496
16	GNMCK28F	17936	18486
16	GNMCK28R	16766	17104
16	GNMCK32F	20788	21317
16	GNMCK32R	21768	22345
16	GNMCK85F	4360	4910
16	GNMCK85R	5620	6191
16	GNMCL06F	5123	5624
16	GNMCL06R	3812	4383
16	GNMCL34F	28058	28532
16	GNMCL34R	26957	27535
16	GNMCL63F	31053	31621
16	GNMCL63R	32284	32700
16	GNMCL70F	26168	26684
16	GNMCM31F	50181	50817
16	GNMCM31R	48867	49582
16	GNMCN28F	69538	70215
16	GNMCN28R	68459	69068
16	GNMCN84F	68423	69040
16	GNMCN84R	66998	67589
16	GNMCO18F	2622	3166
16	GNMCO18R	1677	2332
16	GNMCO35F	70510	71084
16	GNMCO35R	69198	69780
16	GNMCP19F	46453	47147
16	GNMCP19R	48299	48962
16	GNMCP43F	14799	15124
16	GNMCQ02F	19223	19930
16	GNMCQ02R	20338	21001
16	GNMCQ22F	21355	22030
16	GNMCQ22R	19917	20600
16	GNMCQ53F	7175	7907

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
16	GNMCQ53R	8198	8928
16	GNMCQ96R	29546	30182
16	GNMCS41F	29075	29776
16	GNMCS68F	9040	9703
16	GNMCS75R	1277	1893
16	GNMCS76F	2498	3167
16	GNMCV38F	37452	37889
16	GNMCV55R	34048	34804
16	GNMCV60F	59043	59536
16	GNMCV60R	57614	58367
16	GNMCX12F	3746	4302
16	GNMCX12R	5111	5734
16	GNMCX21F	11333	11997
16	GNMCX21R	10200	10848
16	GNMCX63F	225	712
16	GNMCY14F	72030	72750
16	GNMCY14R	70731	71300
16	GNMCY23F	43229	43994
16	GNMCY23R	42063	42641
16	GNMCY41F	27768	28553
16	GNMCY41R	28801	29356
16	GNMCY50F	59253	60030
16	GNMCY50R	58094	58480
16	GNMCY59F	48831	49574
16	GNMCY59R	50018	50543
16	GNMCZ40F	12172	12645
16	GNMCZ40R	13578	14094
16	GNMCZ41F	60265	60795
16	GNMCZ41R	61535	62088
16	GNMCZ80F	29797	30278
16	GNMCZ80R	28542	29086
16	GNMCZ90R	34086	34573
17	GNMAA23F	31103	31553
17	GNMAA23R	32120	32558
17	GNMAA31F	20779	21295
17	GNMAA31R	21615	22086
17	GNMAA67F	32770	33282
17	GNMAA67R	33955	34310
17	GNMAB08F	35151	35717
17	GNMAB08R	33887	34310
17	GNMBA18F	51385	51952
17	GNMBA36F	8398	8967
17	GNMBA36R	9832	10331
17	GNMBA54F	57853	58426
17	GNMBA54R	56651	57182
17	GNMBA74F	22767	23336

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
17	GNMBA74R	21413	21911
17	GNMBA85F	33077	33648
17	GNMBA85R	31797	32251
17	GNMCA19F	36042	36621
17	GNMCB06F	26433	26953
17	GNMCB06R	28247	28714
17	GNMCB10F	38250	38813
17	GNMCB10R	36756	37384
17	GNMCB82F	31729	32377
17	GNMCB82R	32858	33235
17	GNMCF22F	37912	38405
17	GNMCF22R	36753	37421
17	GNMCK05F	7321	7797
17	GNMCK05R	5987	6514
17	GNMCK57F	39678	40046
17	GNMCK57R	40958	41325
17	GNMCM38F	10453	11189
17	GNMCM38R	11737	12393
17	GNMCM58F	22688	23288
17	GNMCM58R	23628	24315
17	GNMCN30F	55573	56235
17	GNMCN30R	56832	57420
17	GNMCO01F	27343	28038
17	GNMCO07F	12194	12723
17	GNMCO07R	13433	14166
17	GNMCO26R	5725	6371
17	GNMCO43F	35750	36434
17	GNMCO43R	37161	37681
17	GNMCO44F	32920	33658
17	GNMCO44R	31733	32327
17	GNMCO55F	10439	11147
17	GNMCO55R	12310	12961
17	GNMCO56F	54670	55322
17	GNMCO56R	55704	56309
17	GNMCP57F	10671	10932
17	GNMCP57R	8680	9034
17	GNMCP66F	57727	58211
17	GNMCP66R	58838	59416
17	GNMCQ42F	22050	22733
17	GNMCQ42R	23218	23942
17	GNMCQ81F	41410	42152
17	GNMCQ81R	42968	43610
17	GNMCS03F	707	1334
17	GNMCS35F	52431	53137
17	GNMCS44F	35071	35764
17	GNMCS70F	6806	7540

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
17	GNMCS89F	38449	39120
17	GNMCS90R	39272	39972
17	GNMCV42F	51980	52438
17	GNMCV92F	43715	44212
17	GNMCV92R	42381	43040
17	GNMCX53F	18076	18436
17	GNMCX53R	16632	17267
17	GNMCY21F	26276	26984
17	GNMCY21R	25220	25785
17	GNMCY43F	55511	56209
17	GNMCY58F	10946	11675
17	GNMCY58R	9574	10130
17	GNMCZ14F	4034	4557
17	GNMCZ14R	5449	5997
17	GNMCZ81F	12505	13016
17	GNMCZ81R	10929	11485
18	GNMAA24F	14784	15300
18	GNMAA24R	15822	16278
18	GNMAA91F	3107	3623
18	GNMAA93F	14115	14633
18	GNMAA93R	12779	13156
18	GNMAB47F	6436	7001
18	GNMCA24F	17599	18212
18	GNMCB51F	10483	11109
18	GNMCB51R	9080	9547
18	GNMCK79F	4421	4931
18	GNMCK79R	5949	6533
18	GNMCM27F	17624	18228
18	GNMCM27R	16432	17178
18	GNMCM56F	13615	14160
18	GNMCM56R	14770	15435
18	GNMCN40R	15893	16523
18	GNMCN44F	14468	15195
18	GNMCN44R	15922	16524
18	GNMCP83F	14201	14738
18	GNMCP83R	15673	16259
18	GNMCY13F	2490	3240
18	GNMCZ03F	14791	15109
18	GNMCZ03R	16087	16657
18	GNMCZ15F	6918	7405
18	GNMCZ15R	5483	6044
18	GNMCZ61F	15232	15736
18	GNMCZ61R	16804	17347
19	GNMAA25F	3689	4210
19	GNMAA25R	4679	5150
19	GNMAA53F	17218	17584

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
19	GNMAA53R	16131	16651
19	GNMAB22F	11317	11854
19	GNMBA56F	29237	29799
19	GNMBB20F	42956	43521
19	GNMBB20R	41743	42275
19	GNMCB08F	1626	2186
19	GNMCB08R	2749	3408
19	GNMCB49F	24542	25193
19	GNMCB49R	23154	23800
19	GNMCB50F	1442	2136
19	GNMCB50R	457	1122
19	GNMCB84F	25574	26173
19	GNMCB84R	24112	24577
19	GNMCD36F	32463	32986
19	GNMCF17F	11187	11695
19	GNMCF17R	9855	10520
19	GNMCF56F	43830	44301
19	GNMCF56R	42446	43137
19	GNMCF62F	46052	46506
19	GNMCH41R	48920	49204
19	GNMCK19F	5471	5977
19	GNMCK19R	6934	7451
19	GNMCK60F	19464	19828
19	GNMCK60R	20624	21189
19	GNMCL07F	29947	30379
19	GNMCL07R	31253	31828
19	GNMCL47F	13187	13681
19	GNMCL47R	11739	12309
19	GNMCL67R	10328	10861
19	GNMCM83F	7074	7667
19	GNMCM83R	5824	6505
19	GNMCM87R	6816	7475
19	GNMCN69F	21718	22367
19	GNMCN69R	23279	23896
19	GNMCO19F	7892	8641
19	GNMCO19R	6509	7230
19	GNMCQ23F	22847	23439
19	GNMCQ23R	24531	25070
19	GNMCQ63F	24578	25176
19	GNMCQ63R	23445	24129
19	GNMCS09F	31343	31944
19	GNMCS34F	32710	33397
19	GNMCV13F	11334	11854
19	GNMCV14R	10046	10690
19	GNMCX15F	8333	9060
19	GNMCX15R	10180	10827

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
19	GNMCX27F	8333	9060
19	GNMCX27R	10188	10827
19	GNMCX56F	40847	41206
19	GNMCX56R	41903	42589
19	GNMCX87F	33938	34084
19	GNMCX87R	31658	32349
19	GNMCY07F	37467	38035
19	GNMCZ04R	24360	24843
20	GNMAA26F	11314	11834
20	GNMAA34R	15825	16187
20	GNMBA46F	9402	9971
20	GNMBA83F	9481	10050
20	GNMBA83R	11039	11224
20	GNMBA92F	3716	4284
20	GNMBA92R	2437	2882
20	GNMCA93F	10570	11228
20	GNMCB42F	12316	12924
20	GNMCB42R	10720	11380
20	GNMCF68F	145	549
20	GNMCS13F	3147	3776
20	GNMCS19F	3135	3707
20	GNMCV43F	4932	5463
20	GNMCV43R	3493	4272
20	GNMCX01R	8929	9576
20	GNMCX32F	2827	3562
20	GNMCX32R	1753	2386
21	GNMAA29F	7970	8459
21	GNMAA29R	6973	7381
21	GNMAA79F	60518	61036
21	GNMAA79R	61382	61783
21	GNMAB13F	91199	91695
21	GNMAB13R	90065	90490
21	GNMAB15F	18098	18666
21	GNMAB15R	17086	17514
21	GNMAB38F	89228	89794
21	GNMAB49F	90018	90554
21	GNMAB53F	57858	58423
21	GNMAB76F	69791	70359
21	GNMAB76R	71099	71621
21	GNMBA08F	88398	88961
21	GNMBA08R	89946	90480
21	GNMBA62F	91149	91717
21	GNMBA62R	90149	90587
21	GNMBB08F	57329	57895
21	GNMBB08R	58629	59155
21	GNMCB36F	86172	86807

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
21	GNMCB36R	87700	88359
21	GNMCB40F	55242	55889
21	GNMCB40R	56581	57269
21	GNMCD13F	26267	26840
21	GNMCD13R	24739	25235
21	GNMCD14F	63282	63678
21	GNMCD22F	39214	39744
21	GNMCD89F	20621	21136
21	GNMCD89R	19243	19626
21	GNMCE04F	48264	48570
21	GNMCE16F	8955	9401
21	GNMCE16R	10419	10933
21	GNMCK72F	28120	28413
21	GNMCK72R	29725	30288
21	GNMCK82F	16224	16679
21	GNMCK82R	17910	18284
21	GNMCK92F	21493	21930
21	GNMCK92R	22899	23382
21	GNMCL15F	15475	16027
21	GNMCL15R	16323	16894
21	GNMCL18F	40761	41272
21	GNMCL18R	39414	39980
21	GNMCL35F	58131	58677
21	GNMCL35R	56683	57252
21	GNMCM02F	77632	78241
21	GNMCM02R	76049	76774
21	GNMCM42F	44749	45453
21	GNMCM51F	70991	71600
21	GNMCM51R	72059	72786
21	GNMCM59F	46177	46805
21	GNMCM59R	47628	48296
21	GNMCM67F	58893	59524
21	GNMCM67R	57080	57810
21	GNMCN01F	29541	30134
21	GNMCN03R	26156	26805
21	GNMCN04F	27776	28333
21	GNMCN07F	3923	4589
21	GNMCN20F	23898	24435
21	GNMCN20R	22616	23262
21	GNMCN38R	27178	27843
21	GNMCN42F	28721	29325
21	GNMCN42R	27182	27579
21	GNMCN48F	31545	32275
21	GNMCN48R	30254	30829
21	GNMCN56F	38871	39524
21	GNMCN56R	37891	38510

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
21	GNMCN74R	76122	76780
21	GNMCN76F	76705	77420
21	GNMCN87F	81602	82287
21	GNMCN87R	80523	81067
21	GNMCO27F	12120	12686
21	GNMCO27R	10881	11591
21	GNMCO37R	5718	6199
21	GNMCO40F	81181	81864
21	GNMCO40R	80087	80668
21	GNMCO41F	64583	65194
21	GNMCO41R	63303	63895
21	GNMCO62F	24786	25412
21	GNMCO62R	23316	23927
21	GNMCO69F	29872	30526
21	GNMCO69R	28732	29361
21	GNMCP53R	42566	43118
21	GNMCP68F	17274	17781
21	GNMCP68R	18590	19166
21	GNMCP78F	20880	21383
21	GNMCP78R	22662	23004
21	GNMCQ50F	52354	53060
21	GNMCQ50R	53094	53813
21	GNMCQ56F	24974	25298
21	GNMCQ56R	26318	26936
21	GNMCQ76F	26247	26921
21	GNMCQ76R	27401	28002
21	GNMCQ86F	45276	45978
21	GNMCQ86R	46636	47364
21	GNMCS08F	7772	7922
21	GNMCS22F	49814	50311
21	GNMCS62F	56147	56850
21	GNMCS82F	1052	1732
21	GNMCW22F	55865	56223
21	GNMCX02R	45344	45988
21	GNMCX09F	6251	6961
21	GNMCX09R	4718	5291
21	GNMCX16F	60624	61395
21	GNMCX16R	59855	60393
21	GNMCX60F	40043	40437
21	GNMCX60R	41031	41715
21	GNMCX74F	59663	60376
21	GNMCX74R	58460	59136
21	GNMCY45F	42419	43108
21	GNMCY45R	44124	44642
21	GNMCY64F	58336	59059
21	GNMCY64R	57045	57582

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
21	GNMCZ28F	82973	83440
21	GNMCZ28R	81697	82250
21	GNMCZ46F	28043	28521
21	GNMCZ46R	26632	27064
21	GNMCZ77F	22158	22671
21	GNMCZ77R	23472	24017
22	GNMAA30F	2165	2683
22	GNMAA30R	3510	3980
22	GNMBA03F	25307	25874
22	GNMCB39F	5720	6103
22	GNMCB39R	3638	3945
22	GNMCK48F	14049	14546
22	GNMCK48R	12667	13251
22	GNMCL28F	17498	18022
22	GNMCL28R	16124	16700
22	GNMCM15R	284	872
22	GNMCN47R	4247	4891
22	GNMCO22F	9932	10637
22	GNMCO22R	11087	11794
22	GNMCO23F	10489	11080
22	GNMCO23R	11662	12303
22	GNMCQ04F	25363	26023
22	GNMCQ04R	24009	24693
22	GNMCS17F	5636	6187
22	GNMCS20F	21715	22271
22	GNMCV45F	11101	11552
22	GNMCV45R	12185	12992
22	GNMCV65F	21938	22388
22	GNMCW11F	21268	21882
22	GNMCZ08F	9245	9752
22	GNMCZ56R	4001	4481
22	GNMCZ57F	92	610
22	GNMCZ57R	1391	1949
23	GNMAA32R	501	916
24	GNMAA32F	34126	34644
24	GNMAA78F	12905	13389
24	GNMAA78R	11993	12173
24	GNMAA92F	5430	5906
24	GNMAA92R	6781	6979
24	GNMBA28F	25580	26147
24	GNMBA28R	24581	24744
24	GNMBA64F	44750	45281
24	GNMBA64R	43715	43924
24	GNMCA03F	47978	48229
24	GNMCA11F	5227	5845
24	GNMCB53F	31273	31860

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
24	GNMCB53R	29940	30477
24	GNMCD60F	49318	49836
24	GNMCF28R	25897	26427
24	GNMCF33F	53794	54122
24	GNMCF33R	55250	55649
24	GNMCF55F	18332	18818
24	GNMCF55R	16670	17304
24	GNMCF88F	31085	31484
24	GNMCF88R	29803	30387
24	GNMCF94F	32330	32765
24	GNMCF94R	30474	31147
24	GNMCH39F	20653	21054
24	GNMCH71F	20501	20708
24	GNMCK74F	31152	31629
24	GNMCK74R	32456	33004
24	GNMCK94F	19578	20116
24	GNMCK94R	18366	18866
24	GNMCL74F	16135	16693
24	GNMCL74R	18346	18913
24	GNMCM07F	48543	49161
24	GNMCM07R	47427	48064
24	GNMCM72F	14897	15471
24	GNMCM72R	15789	16445
24	GNMCM86F	32288	32811
24	GNMCM86R	31171	31832
24	GNMCN14F	11430	12112
24	GNMCN14R	12286	12980
24	GNMCN59F	46864	47475
24	GNMCN59R	47935	48525
24	GNMCN60F	22771	23206
24	GNMCN60R	24286	24873
24	GNMCN91F	1694	2415
24	GNMCN91R	411	1022
24	GNMCO65F	4379	5044
24	GNMCO65R	5399	6070
24	GNMCO91F	54004	54574
24	GNMCO91R	55258	55836
24	GNMCP23F	21885	22586
24	GNMCP23R	20351	20912
24	GNMCP71F	53062	53612
24	GNMCP71R	54382	54958
24	GNMCQ33F	31360	32059
24	GNMCQ33R	30167	30816
24	GNMCS10F	52384	52999
24	GNMCS79R	9557	10245
24	GNMCV21F	13147	13602

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
24	GNMCV22R	14356	15028
24	GNMCV63F	11801	12250
24	GNMCV63R	12681	13494
24	GNMCV66F	53565	54040
24	GNMCV66R	52285	53073
24	GNMCV73R	42644	43443
24	GNMCV78F	23665	24161
24	GNMCV78R	24559	25362
24	GNMCX22F	8574	9293
24	GNMCX22R	9681	10320
24	GNMCX33F	23234	23994
24	GNMCX33R	21803	22176
24	GNMCX34F	23296	23994
24	GNMCX34R	21787	22355
24	GNMCX40F	28130	28866
24	GNMCX40R	29005	29697
24	GNMCX70F	10118	10635
24	GNMCX70R	11461	12043
24	GNMCX72F	27541	27741
24	GNMCY35F	32221	32765
24	GNMCY35R	31087	31546
24	GNMCY55F	45603	46359
24	GNMCY66R	2897	3449
24	GNMCY77F	29179	29866
24	GNMCY77R	27766	28254
24	GNMCY82F	9582	10184
24	GNMCY82R	11010	11421
24	GNMCY94F	6998	7520
24	GNMCY96F	22341	22994
24	GNMCY96R	23886	24294
24	GNMCZ37F	24346	24873
24	GNMCZ37R	23379	23953
25	GNMAA34F	450	701
25	GNMBA48F	4952	5519
25	GNMBA48R	4021	4222
25	GNMCA16F	14824	15438
25	GNMCB09F	22420	22990
25	GNMCB09R	23872	24453
25	GNMCD04F	2415	2961
25	GNMCD04R	1176	1633
25	GNMCK09F	3101	3667
25	GNMCK09R	4706	5009
25	GNMCK50F	8704	9235
25	GNMCK50R	10150	10511
25	GNMCM76R	3069	3807
25	GNMCM96F	13743	14447

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
25	GNMCM96R	12253	12967
25	GNMCN04R	15105	15705
25	GNMCN05F	13789	14465
25	GNMCP16F	9455	10151
25	GNMCP16R	8452	9076
25	GNMCP62R	9951	10498
25	GNMCX61F	2026	2420
25	GNMCX61R	3150	3850
25	GNMCY04F	10646	11249
25	GNMCY04R	12076	12645
25	GNMCZ20F	13438	13952
25	GNMCZ20R	12311	12861
26	GNMAA37F	45118	45485
26	GNMAA37R	46181	46702
26	GNMAA44F	38832	39198
26	GNMAA44R	37468	37990
26	GNMBB25F	2584	3149
26	GNMBB25R	4308	4852
26	GNMCA28F	34335	34909
26	GNMCB61F	37090	37496
26	GNMCE76F	146	542
26	GNMCE76R	1633	1980
26	GNMCF66F	27879	28279
26	GNMCF66R	29423	30059
26	GNMCL21F	39439	39981
26	GNMCL21R	37698	38064
26	GNMCL69F	3546	4121
26	GNMCL69R	4207	4797
26	GNMCM34R	3940	4653
26	GNMCM89F	5891	6343
26	GNMCM89R	7010	7718
26	GNMCM92R	30750	31399
26	GNMCN54F	28683	29364
26	GNMCN54R	27207	27807
26	GNMCN79F	51540	52223
26	GNMCN79R	50402	50941
26	GNMCO14F	33740	34469
26	GNMCO14R	35347	36067
26	GNMCQ26F	47379	47982
26	GNMCQ26R	48736	49406
26	GNMCS81F	36588	37281
26	GNMCS88F	19142	19409
26	GNMCS89R	17251	18014
26	GNMCV32F	18068	18514
26	GNMCV33F	30470	30781
26	GNMCV33R	28683	29309

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
26	GNMCV70F	41545	42025
26	GNMCV70R	42579	43282
26	GNMCV76F	30234	30720
26	GNMCV76R	31359	32063
26	GNMCV86R	42591	43300
26	GNMCV87F	41330	41805
26	GNMCV87R	42509	43300
26	GNMCX26R	42058	42510
26	GNMCY31R	1275	1860
26	GNMCY86F	27767	28402
26	GNMCY86R	26306	26736
26	GNMCZ13F	23798	24317
26	GNMCZ13R	24994	25572
26	GNMCZ64F	26763	27169
26	GNMCZ64R	27996	28534
26	GNMCZ71F	47451	47955
26	GNMCZ71R	46061	46606
26	GNMCZ95R	8013	8499
26	GNMCZ96R	8005	8483
27	GNMAA41F	3036	3402
27	GNMAA41R	2156	2677
27	GNMAA65F	58776	59296
27	GNMAA65R	60307	60457
27	GNMAB83F	38177	38746
27	GNMAB83R	36806	37326
27	GNMAB86F	20818	21390
27	GNMAB86R	21914	22429
27	GNMAB92F	21743	22226
27	GNMBA25F	28880	29408
27	GNMBA25R	27506	28043
27	GNMBA49F	40184	40752
27	GNMCB28F	15988	16497
27	GNMCB28R	14642	15180
27	GNMCB30R	14648	14996
27	GNMCB35F	33768	34099
27	GNMCB35R	32048	32548
27	GNMCB37F	31837	32567
27	GNMCB37R	30832	31421
27	GNMCB58F	30329	31041
27	GNMCB58R	31809	32460
27	GNMCD63F	15824	16290
27	GNMCD79F	63644	64156
27	GNMCD79R	62110	62364
27	GNMCF64F	41517	41871
27	GNMCF84F	518	956
27	GNMCF84R	1834	2533

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
27	GNMCF85F	6358	6815
27	GNMCF85R	7660	8383
27	GNMCH76F	22610	22966
27	GNMCH77F	22613	22953
27	GNMCK01F	62394	62733
27	GNMCK01R	60888	61415
27	GNMCK18F	66502	66997
27	GNMCK18R	65282	65724
27	GNMCK25F	27644	28213
27	GNMCK61F	32761	33107
27	GNMCK61R	30995	31329
27	GNMCK76F	19006	19542
27	GNMCK76R	17573	18122
27	GNMCK81F	61093	61511
27	GNMCK81R	59863	60445
27	GNMCK87F	36665	36996
27	GNMCK87R	34928	35498
27	GNMCL44F	38519	39001
27	GNMCL44R	37283	37863
27	GNMCL76F	49805	50300
27	GNMCL76R	48285	48854
27	GNMCM23F	27097	27789
27	GNMCM23R	25771	26483
27	GNMCN12F	8559	9239
27	GNMCN12R	7161	7752
27	GNMCN13F	68144	68833
27	GNMCN13R	66871	67394
27	GNMCN17F	36140	36815
27	GNMCN17R	35179	35753
27	GNMCN18F	55803	56468
27	GNMCN18R	54618	55229
27	GNMCN34F	59534	60268
27	GNMCN34R	19457	20056
27	GNMCN38F	17990	18719
27	GNMCN61F	18037	18594
27	GNMCN61R	19452	20056
27	GNMCN70F	32750	33421
27	GNMCN80R	37432	38115
27	GNMCN81F	38597	39329
27	GNMCN81R	37434	38096
27	GNMCO02R	59813	60549
27	GNMCO38F	51253	51930
27	GNMCO52R	33701	34400
27	GNMCO57F	37843	38469
27	GNMCO57R	36757	37320
27	GNMCP50F	7088	7522

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
27	GNMCP50R	5679	6058
27	GNMCQ93R	2933	3510
27	GNMCS49F	11768	12343
27	GNMCV50F	28795	29193
27	GNMCV50R	27644	28413
27	GNMCV85F	21568	22089
27	GNMCV85R	22559	23351
27	GNMCW02F	47088	47658
27	GNMCW24F	56091	56713
27	GNMCY27R	5455	5536
27	GNMCY33F	37884	38598
27	GNMCY33R	39134	39678
27	GNMCY62F	39794	40529
27	GNMCY62R	41156	41683
27	GNMCY63F	39843	40316
27	GNMCY72F	15711	16330
27	GNMCY72R	14681	15239
28	GNMAA45F	4450	4816
28	GNMAA54R	4273	4733
28	GNMCD82F	1790	2266
28	GNMCD82R	3389	3826
28	GNMCO78F	6645	7293
28	GNMCO86F	6688	7310
28	GNMCO86R	8039	8651
28	GNMCW05F	6711	7331
28	GNMCZ09F	13148	13623
28	GNMCZ09R	11925	12279
29	GNMAA47F	27107	27473
29	GNMAA47R	25852	26322
29	GNMAA71F	19984	20503
29	GNMAA71R	21408	21826
29	GNMAA80R	20918	21282
29	GNMAB31F	32769	33333
29	GNMAB31R	31525	31942
29	GNMAB77F	21439	22007
29	GNMAB77R	22335	22857
29	GNMCA22F	9411	10028
29	GNMCB74F	26713	27450
29	GNMCB74R	25839	26476
29	GNMCD08F	17015	17514
29	GNMCD31F	19776	20146
29	GNMCF43F	26320	26631
29	GNMCF43R	27361	28023
29	GNMCF87F	30819	31269
29	GNMCF87R	32125	32845
29	GNMCH41F	30939	31379

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
29	GNMCK20F	2703	3104
29	GNMCK20R	4020	4346
29	GNMCL02F	32166	32619
29	GNMCL02R	33533	33884
29	GNMCL12F	360	831
29	GNMCL12R	1490	2039
29	GNMCL73R	32923	33504
29	GNMCL85R	10861	11425
29	GNMCM77F	17717	18313
29	GNMCM77R	16440	17172
29	GNMCN64F	6192	6750
29	GNMCN64R	7430	8018
29	GNMCN68F	30002	30712
29	GNMCN83F	34059	34776
29	GNMCN83R	32873	33458
29	GNMCO28F	7197	7872
29	GNMCO28R	8396	9089
29	GNMCO53F	20633	21342
29	GNMCO53R	22061	22663
29	GNMCO67F	1523	2102
29	GNMCO67R	2871	3524
29	GNMCP82F	30881	31419
29	GNMCP82R	29550	30117
29	GNMCS26F	30683	31168
29	GNMCS90F	16067	16703
29	GNMCS91R	16949	17757
29	GNMCW09F	3770	4381
29	GNMCY19F	14037	14742
29	GNMCY89F	7491	8173
30	GNMAA48R	1027	1347
30	GNMAB21R	3808	4233
30	GNMCC90F	7658	8102
30	GNMCL10F	2942	3470
30	GNMCL10R	4319	4883
30	GNMCM64R	7645	8319
30	GNMCO63F	12259	12933
30	GNMCO63R	11104	11789
30	GNMCP58F	8513	9047
30	GNMCP58R	10322	10707
30	GNMCV03F	10383	10724
30	GNMCV04R	8992	9749
30	GNMCX06F	11346	12072
30	GNMCX06R	12784	13418
30	GNMCX18F	11968	12726
30	GNMCX18R	13547	14189
30	GNMCX71F	9073	9653

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
30	GNMCX71R	7669	8353
30	GNMCY15F	3214	3933
30	GNMCY15R	1508	2079
31	GNMAA49F	7079	7444
31	GNMAA49R	5736	6260
31	GNMBA38F	692	1262
31	GNMBA79F	7797	8367
31	GNMCL32F	3721	4184
31	GNMCL32R	2230	2815
31	GNMCN88F	1761	2482
31	GNMCN88R	3292	3892
31	GNMCQ51F	3265	3909
31	GNMCQ51R	4295	5012
31	GNMCX63R	7311	8010
31	GNMCY61R	4386	4868
31	GNMCY91F	2862	3456
32	GNMAA52F	1739	2107
32	GNMAA52R	2617	3138
32	GNMAA89F	13148	13666
32	GNMAB90F	5624	6192
32	GNMAB90R	6600	7118
32	GNMCF38F	3403	3878
32	GNMCF38R	4584	5237
32	GNMCK38F	6598	7143
32	GNMCK38R	5207	5792
32	GNMCP85F	6949	7473
32	GNMCP85R	5282	5869
32	GNMCQ07F	10995	11623
32	GNMCQ07R	12678	13358
32	GNMCV23F	5455	5912
32	GNMCV24R	4006	4751
32	GNMCX13F	9897	10671
32	GNMCX13R	8710	9345
32	GNMCX45F	3857	4557
32	GNMCX45R	2724	3424
32	GNMCX57F	6426	6642
32	GNMCX57R	6424	6642
32	GNMCY06F	10183	10812
32	GNMCY06R	9259	9808
33	GNMAA57F	2954	3324
33	GNMAA57R	1924	2445
33	GNMAB30F	5838	6402
33	GNMAB30R	4864	5193
33	GNMAB48F	8816	9381
33	GNMBA50F	7809	8374
33	GNMBA50R	6161	6686

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
33	GNMCA25F	18305	18918
33	GNMCA80F	3189	3849
33	GNMCL88F	12941	13492
33	GNMCL88R	11494	12068
33	GNMCM57F	6934	7569
33	GNMCM57R	7814	8548
33	GNMCN49F	18067	18780
33	GNMCN49R	16729	17352
33	GNMCO54F	17815	18524
33	GNMCO54R	16974	17598
33	GNMCP59F	13173	13661
33	GNMCP59R	14688	15102
33	GNMCQ29F	13338	14036
33	GNMCQ29R	11998	12686
33	GNMCQ87F	5967	6647
33	GNMCQ87R	7354	7981
33	GNMCS47F	7736	8461
33	GNMCV30F	18040	18529
33	GNMCV31F	1808	2296
33	GNMCV31R	16473	17092
33	GNMCV32R	2897	3643
33	GNMCY12F	13632	14327
33	GNMCY12R	14891	15465
33	GNMCZ12F	14374	14860
33	GNMCZ12R	12879	13414
34	GNMAA59R	20271	20600
34	GNMAB63F	21594	22082
34	GNMAB87F	4234	4656
34	GNMAB93F	8137	8678
34	GNMAB93R	7021	7543
34	GNMBA26F	17728	18076
34	GNMBA31R	20426	20952
34	GNMBA60F	2998	3562
34	GNMBA60R	4887	5305
34	GNMBA89F	12688	13184
34	GNMBA89R	11336	11869
34	GNMBA90F	1963	2532
34	GNMBA90R	3410	3918
34	GNMBB10F	18931	19469
34	GNMBB10R	20494	20791
34	GNMCA73F	10776	11434
34	GNMCD09F	1576	2151
34	GNMCD09R	202	580
34	GNMCL40F	6504	7032
34	GNMCL40R	7906	8476
34	GNMCM41F	15257	15722

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
34	GNMCM41R	13646	14279
34	GNMCM84F	10143	10755
34	GNMCM84R	11418	12090
34	GNMCP65R	13124	13566
34	GNMCQ57F	1107	1637
34	GNMCQ57R	2550	3230
34	GNMCV15F	10810	11260
34	GNMCV16R	9522	10243
34	GNMCX35F	24683	25380
34	GNMCX35R	25964	26651
34	GNMCX48F	27078	27683
34	GNMCX48R	25636	26324
34	GNMCZ82R	4431	4970
35	GNMAA60R	9724	9928
35	GNMAA81R	42064	42495
35	GNMAB09F	29605	30171
35	GNMBA37F	1865	2426
35	GNMBA37R	755	1265
35	GNMCA66F	14095	14490
35	GNMCB95F	29548	30210
35	GNMCB95R	28364	28994
35	GNMCD41F	4298	4824
35	GNMCD41R	2960	3326
35	GNMCD49F	47011	47510
35	GNMCD49R	45671	46032
35	GNMCD52F	46968	47374
35	GNMCE13F	44763	45068
35	GNMCE13R	43656	44020
35	GNMCK86F	32959	33472
35	GNMCL94F	45671	46185
35	GNMCL94R	44388	44948
35	GNMCM08F	32206	32865
35	GNMCM08R	33769	34324
35	GNMCN16F	11716	12326
35	GNMCN16R	10117	10693
35	GNMCN33F	2863	3568
35	GNMCN33R	4337	4927
35	GNMCO11F	117	667
35	GNMCO11R	1479	2220
35	GNMCO20F	41254	41858
35	GNMCO20R	42840	43385
35	GNMCP03R	15135	15820
35	GNMCP33F	33871	34386
35	GNMCP33R	31902	32446
35	GNMCS31F	25024	25611
35	GNMCS80F	26013	26719

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
35	GNMCV20F	11142	11598
35	GNMCV21R	9547	10242
35	GNMCV41F	1508	1764
35	GNMCV41R	2993	3375
35	GNMCV46F	19148	19638
35	GNMCX37F	10287	10978
35	GNMCX75F	16758	17496
35	GNMCX75R	17915	18615
35	GNMCY38F	35286	36002
35	GNMCY38R	36447	37009
35	GNMCZ63F	17628	18139
35	GNMCZ63R	16308	16866
36	GNMAA61F	17639	18003
36	GNMAA61R	19148	19669
36	GNMAB14F	9325	9894
36	GNMAB14R	10480	10900
36	GNMAB23F	5098	5510
36	GNMAB23R	5999	6420
36	GNMBA04F	7545	8114
36	GNMBA04R	8552	9087
36	GNMCB81F	1908	2616
36	GNMCB81R	1189	1739
36	GNMCD86F	266	753
36	GNMCD86R	1917	2276
36	GNMCL29F	19188	19732
36	GNMCL46F	5977	6459
36	GNMCL46R	6855	7431
36	GNMCL71R	2286	2862
36	GNMCN74F	8750	9460
36	GNMCN76R	7557	8138
36	GNMCP37R	5055	5645
36	GNMCS39F	3380	4120
36	GNMCV57F	6730	7217
36	GNMCV57R	7760	8463
36	GNMCX54F	7658	7977
36	GNMCX54R	6197	6884
36	GNMCY85R	6699	7077
36	GNMCZ06F	17782	18302
36	GNMCZ73F	15242	15755
37	GNMAA64F	11674	12041
37	GNMAA64R	10619	11088
37	GNMAB25F	25946	26508
37	GNMAB25R	27013	27437
37	GNMAB32R	446	844
37	GNMAB89F	2515	3085
37	GNMAB89R	3403	3923

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
37	GNMAB91F	19524	19900
37	GNMAB91R	18389	18909
37	GNMCA84F	8986	9651
37	GNMCA92F	10174	10831
37	GNMCB13F	28388	28959
37	GNMCB44F	17203	17885
37	GNMCB44R	16050	16676
37	GNMCB72F	15012	15708
37	GNMCB72R	16365	16857
37	GNMCD32F	4633	5112
37	GNMCD32R	2775	3142
37	GNMCD34F	21613	22123
37	GNMCD34R	23152	23452
37	GNMCD43F	23745	24277
37	GNMCF03F	23267	23766
37	GNMCF03R	21815	22457
37	GNMCK16F	12575	13127
37	GNMCK69R	981	1281
37	GNMCL41F	4846	5357
37	GNMCL41R	6380	6932
37	GNMCM06R	17272	17986
37	GNMCM82F	14731	15358
37	GNMCM82R	15814	16507
37	GNMCQ08F	20211	20740
37	GNMCQ08R	18866	19521
37	GNMCQ59F	16099	16826
37	GNMCQ59R	15132	15853
37	GNMCS58F	16358	17054
37	GNMCV94F	21841	22327
37	GNMCV94R	20477	21267
37	GNMCX07F	25522	26245
37	GNMCX07R	26310	26960
37	GNMCX69F	10320	10866
37	GNMCX69R	11842	12449
37	GNMCX93F	7947	8360
37	GNMCX93R	6445	6970
37	GNMCY18F	10778	11193
37	GNMCY18R	9630	10203
37	GNMCY67F	26216	26689
37	GNMCY67R	24586	24992
37	GNMCZ87F	28035	28543
37	GNMCZ87R	26386	26930
38	GNMAA74F	185	702
38	GNMAB59F	370	710
38	GNMCM68F	512	991
39	GNMBA35F	3187	3756

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
39	GNMCL49F	518	1006
39	GNMCM19F	3839	4413
39	GNMCM19R	2735	3480
39	GNMCM68R	3717	4374
39	GNMCN15F	11	695
39	GNMCN15R	1589	2036
39	GNMCS14F	2485	3018
39	GNMCV29F	4010	4481
39	GNMCV30R	2621	3321
39	GNMCZ91F	4347	4839
39	GNMCZ91R	3070	3594
40	GNMAA75F	1493	2009
40	GNMBA84F	14749	15315
40	GNMBA84R	13039	13401
40	GNMBB27F	7061	7629
40	GNMBB27R	5877	6280
40	GNMCA65F	10805	11468
40	GNMCF01F	9566	10068
40	GNMCF01R	7689	8249
40	GNMCF52F	13446	13800
40	GNMCF52R	14807	15448
40	GNMCK41F	1322	1894
40	GNMCK41R	1	549
40	GNMCN01R	8094	8669
40	GNMCN02F	6573	7152
40	GNMCY39F	12214	12932
40	GNMCY39R	11377	11773
40	GNMCZ75F	4573	5040
40	GNMCZ75R	3272	3824
41	GNMAA82F	1944	2123
41	GNMAA82R	540	848
41	GNMCA09F	4155	4769
41	GNMCL45F	5831	6382
41	GNMCL45R	7014	7592
41	GNMCX84F	6407	7029
41	GNMCX84R	4937	5630
41	GNMCZ07F	753	1256
41	GNMCZ07R	2139	2681
42	GNMAA85F	33488	34005
42	GNMAA85R	34461	34906
42	GNMAB11F	27021	27587
42	GNMAB16F	16195	16762
42	GNMAB16R	17262	17683
42	GNMAB51F	32336	32901
42	GNMAB64F	9048	9478
42	GNMBA52F	25714	26279

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
42	GNMBA52R	26930	27429
42	GNMBA63F	25856	26418
42	GNMCA10F	9199	9803
42	GNMCA90F	12306	12957
42	GNMCD76F	43170	43607
42	GNMCD80F	25485	25983
42	GNMCD80R	24100	24472
42	GNMCD81F	25467	25981
42	GNMCF21F	42792	43250
42	GNMCF21R	43820	44488
42	GNMCF79F	19953	20412
42	GNMCF79R	18429	19107
42	GNMCH08F	10638	10983
42	GNMCH61F	35608	36017
42	GNMCK58F	11541	12006
42	GNMCK58R	13419	13981
42	GNMCM03R	37448	38182
42	GNMCM48F	1	622
42	GNMCM48R	1215	1878
42	GNMCO34F	11655	12379
42	GNMCO34R	10537	11201
42	GNMCO70R	39192	39848
42	GNMCO84F	24768	25509
42	GNMCO84R	24098	24770
42	GNMCP29F	40509	41019
42	GNMCP29R	38958	39359
42	GNMCQ60F	38032	38565
42	GNMCQ69F	8563	9122
42	GNMCQ69R	6981	7666
42	GNMCS69F	3213	3921
42	GNMCV25F	17625	18095
42	GNMCV26R	16021	16633
42	GNMCX46F	4775	5450
42	GNMCX46R	3438	4125
42	GNMCX88R	17104	17778
42	GNMCY37F	7223	7838
42	GNMCY37R	5827	6323
42	GNMCY69F	22213	22853
42	GNMCY69R	21279	21796
42	GNMCZ85F	19300	19813
43	GNMAA86F	5244	5760
43	GNMAA86R	4311	4783
43	GNMCS54F	3163	3797
43	GNMCV84F	1109	1600
43	GNMCV84R	2002	2781
44	GNMAA87F	26931	27447

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
44	GNMAA87R	27952	28361
44	GNMAA90F	6714	7230
44	GNMAA90R	8124	8276
44	GNMAB27F	4036	4606
44	GNMAB27R	4904	5327
44	GNMCD11F	4246	4813
44	GNMCD11R	5623	6146
44	GNMCQ17F	6327	7009
44	GNMCQ17R	7631	8317
44	GNMCQ67F	1410	2013
44	GNMCQ67R	2571	3261
44	GNMCS92F	21392	22037
44	GNMCS94R	22779	23479
44	GNMCS96F	22613	22986
44	GNMCX79F	14815	15344
44	GNMCX79R	16086	16760
44	GNMCZ44F	19312	19820
44	GNMCZ44R	20486	21049
45	GNMAA88F	3827	4313
45	GNMBA05F	7835	8403
45	GNMBA05R	6395	6824
45	GNMCZ39F	143	619
45	GNMCZ39R	1545	2114
46	GNMAA94F	5740	6254
46	GNMAA94R	6575	7044
46	GNMAB29F	659	1225
46	GNMAB29R	1871	2298
46	GNMAB78F	16523	16951
46	GNMAB78R	15145	15666
46	GNMCA05F	4467	5137
46	GNMCD25F	11261	11830
46	GNMCD25R	10056	10529
46	GNMCD45F	4725	5273
46	GNMCD45R	3455	3826
46	GNMCD72F	12772	13251
46	GNMCD72R	14201	14542
46	GNMCK45F	6690	7258
46	GNMCK45R	5280	5857
46	GNMCK53F	9263	9636
46	GNMCK53R	10581	11122
46	GNMCN62R	20059	20606
46	GNMCO72F	11911	12654
46	GNMCO72R	10592	11291
46	GNMCS38F	8266	8953
46	GNMCS50F	9604	10313
46	GNMCY09F	18777	19443

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
46	GNMCY09R	20339	20885
46	GNMCY48F	13317	14054
46	GNMCY48R	12373	12900
47	GNMAB03F	39285	39849
47	GNMAB03R	40395	40825
47	GNMAB57F	8125	8631
47	GNMAB62F	5129	5697
47	GNMAB72F	25957	26522
47	GNMAB72R	26812	27332
47	GNMBA39F	10581	11112
47	GNMBA39R	9272	9805
47	GNMBA68F	33182	33747
47	GNMBA68R	32098	32634
47	GNMBB31F	46909	47485
47	GNMBB31R	45477	45996
47	GNMCB64F	8634	9225
47	GNMCB64R	9880	10466
47	GNMCD39F	26389	26882
47	GNMCF18F	42096	42592
47	GNMCF18R	40473	41111
47	GNMCF47F	46147	46634
47	GNMCF47R	44893	45560
47	GNMCK29F	14259	14820
47	GNMCK29R	12913	13476
47	GNMCK33F	11732	12246
47	GNMCK33R	10377	10759
47	GNMCK51F	19259	19619
47	GNMCK51R	17899	18248
47	GNMCL24F	21022	21491
47	GNMCL24R	19374	19922
47	GNMCL66F	34263	34768
47	GNMCL66R	35478	36049
47	GNMCM30R	35959	36642
47	GNMCM37R	18280	18787
47	GNMCN36F	28250	28958
47	GNMCN73F	29393	30074
47	GNMCN73R	28267	28921
47	GNMCN93F	1262	1971
47	GNMCN93R	2446	2878
47	GNMCO45F	14719	15397
47	GNMCO45R	15952	16635
47	GNMCO49F	38118	38828
47	GNMCO49R	39315	39845
47	GNMCO60F	21461	22152
47	GNMCO60R	19964	20648
47	GNMCO83R	16405	17063

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
47	GNMCP08F	4600	5318
47	GNMCP08R	5704	6436
47	GNMCP12F	44482	45180
47	GNMCP12R	43247	43929
47	GNMCQ70F	28264	28919
47	GNMCQ70R	27232	27902
47	GNMCS79F	28111	28860
47	GNMCX52F	44094	44441
47	GNMCX52R	45425	46100
47	GNMCX73F	8582	9157
47	GNMCX73R	7456	8141
47	GNMCY08F	22073	22785
47	GNMCY08R	20965	21539
47	GNMCY17F	13457	14071
47	GNMCY17R	12199	12710
47	GNMCY60F	4726	5396
47	GNMCY60R	3394	3937
47	GNMCZ72F	26112	26584
47	GNMCZ72R	27111	27642
48	GNMAB10F	45864	46429
48	GNMAB10R	46823	47246
48	GNMAB26F	18205	18771
48	GNMAB26R	17068	17496
48	GNMAB46F	39600	40166
48	GNMAB71F	36266	36835
48	GNMAB71R	35583	35981
48	GNMBA10F	24081	24641
48	GNMBA10R	25627	26158
48	GNMCA01F	2669	3310
48	GNMCA69F	24907	25573
48	GNMCA77F	44240	44904
48	GNMCB68F	48529	49183
48	GNMCB68R	49751	50229
48	GNMCD05F	61093	61524
48	GNMCD05R	47029	47548
48	GNMCD24F	41436	41982
48	GNMCD24R	42664	43161
48	GNMCD70F	43366	43798
48	GNMCE06F	45703	46081
48	GNMCE07R	46605	47129
48	GNMCE90F	6380	6925
48	GNMCE90R	5283	5799
48	GNMCF24F	56963	57448
48	GNMCF24R	55581	56243
48	GNMCF70R	50946	51263
48	GNMCF93F	46705	47157

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
48	GNMCF93R	48122	48692
48	GNMCH40F	24168	24458
48	GNMCH64F	60688	61022
48	GNMCK08F	12988	13530
48	GNMCK08R	11548	12144
48	GNMCK31F	48379	48939
48	GNMCK31R	47177	47731
48	GNMCK46F	13297	13814
48	GNMCK46R	12071	12654
48	GNMCK56F	29433	29963
48	GNMCK56R	27927	28487
48	GNMCK70F	41792	42156
48	GNMCK70R	43324	43888
48	GNMCL05F	22552	23041
48	GNMCL05R	21742	22293
48	GNMCL61F	15321	15724
48	GNMCL61R	14006	14449
48	GNMCL86F	23803	24358
48	GNMCL86R	22389	22965
48	GNMCM40F	60172	60784
48	GNMCM40R	43992	44623
48	GNMCM49R	63033	63741
48	GNMCM60F	28595	29249
48	GNMCM60R	27285	27929
48	GNMCN95F	21768	22424
48	GNMCN96F	52482	53159
48	GNMCO12F	49771	50550
48	GNMCO12R	49060	49698
48	GNMCO76F	26934	27624
48	GNMCO76R	25392	26062
48	GNMCO90F	10121	10652
48	GNMCO90R	8744	9318
48	GNMCP81F	26207	26575
48	GNMCP81R	27441	28017
48	GNMCQ16R	1	661
48	GNMCQ36R	13779	14476
48	GNMCQ48F	44157	44770
48	GNMCQ48R	43032	43754
48	GNMCQ64F	12475	13200
48	GNMCQ66F	12668	13370
48	GNMCQ66R	13747	14472
48	GNMCQ89F	16922	17619
48	GNMCV06F	48695	49152
48	GNMCV07R	47510	48231
48	GNMCV11F	26723	27238
48	GNMCV12R	27836	28452

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
48	GNMCV18F	35744	36244
48	GNMCV19R	34456	35205
48	GNMCV82F	8278	8644
48	GNMCV82R	8280	8645
48	GNMCV96F	45990	46492
48	GNMCV96R	44480	45162
48	GNMCX28F	42946	43632
48	GNMCX28R	44129	44767
48	GNMCX29F	59233	59998
48	GNMCX29R	58344	58984
48	GNMCX42F	22170	22862
48	GNMCX42R	23577	24264
48	GNMCX50F	29838	30232
48	GNMCX50R	30956	31633
48	GNMCX80F	30061	30735
48	GNMCX80R	31536	32224
48	GNMCY70F	13009	13629
48	GNMCY70R	11725	12281
48	GNMCZ84R	4001	4533
49	GNMAB32F	401	684
50	GNMAB35F	17857	18274
50	GNMBA70F	14615	15180
50	GNMBA70R	15849	16383
50	GNMCB20R	20852	21453
50	GNMCB89F	12569	13223
50	GNMCB89R	14045	14508
50	GNMCF67F	4524	4879
50	GNMCF67R	3257	3858
50	GNMCH89F	19690	20140
50	GNMCH89R	18248	18535
50	GNMCK49F	20201	20665
50	GNMCK49R	18771	19297
50	GNMCM01F	2158	2770
50	GNMCM01R	708	1314
50	GNMCN41F	21893	22570
50	GNMCN41R	23128	23476
50	GNMCO04F	2174	2638
50	GNMCO04R	837	1541
50	GNMCO82F	16481	17139
50	GNMCO82R	17538	18219
50	GNMCP61F	13046	13330
50	GNMCP61R	14605	15154
50	GNMCS33F	27679	28393
50	GNMCV22F	21920	22410
50	GNMCV23R	20644	21369
50	GNMCV47F	17147	17659

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
50	GNMCV47R	18206	18900
50	GNMCV58R	1132	1905
50	GNMCV59R	1242	1814
50	GNMCX41F	3977	4725
50	GNMCX41R	5212	5916
50	GNMCY22F	22454	23103
50	GNMCY29R	22461	23008
50	GNMCY71F	10076	10646
50	GNMCY71R	9041	9543
50	GNMCZ52F	20698	21140
50	GNMCZ52R	22156	22569
50	GNMCZ94F	3890	4317
50	GNMCZ94R	5230	5743
50	GNMCZ95F	3902	4346
50	GNMCZ96F	3902	4346
51	GNMAB39F	5946	6511
51	GNMBA51F	8613	9139
51	GNMBA51R	6844	7329
51	GNMCL84F	7136	7509
51	GNMCL84R	8501	9072
51	GNMCO08R	979	1711
51	GNMCY10F	1194	1921
51	GNMCY10R	50	610
51	GNMCZ33F	3405	3947
51	GNMCZ33R	4668	5244
52	GNMAB40F	15814	16385
52	GNMCB93F	7437	8109
52	GNMCB93R	8732	9304
52	GNMCF69F	9103	9470
52	GNMCF69R	7871	8573
52	GNMCF92F	2901	3235
52	GNMCF92R	1359	2018
52	GNMCL51F	16830	17360
52	GNMCL51R	18234	18580
52	GNMCM61R	12794	13378
52	GNMCN24F	1	676
52	GNMCN24R	1452	2016
52	GNMCO31F	17039	17664
52	GNMCO31R	18187	18861
52	GNMCS05F	11540	12169
52	GNMCX49F	10221	10402
52	GNMCX49R	8569	9260
52	GNMCX96R	4202	4835
52	GNMCZ83F	11839	12349
52	GNMCZ83R	13065	13609
53	GNMAB50F	81	306

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
54	GNMAB60F	4573	5141
54	GNMCD66F	258	750
55	GNMAB66F	1314	1623
55	GNMCB73F	3597	4316
55	GNMCB73R	5062	5644
55	GNMCM35F	3120	3883
55	GNMCM35R	2555	3288
55	GNMCX47F	5496	6201
55	GNMCX47R	4289	4982
55	GNMCY34F	5585	6305
56	GNMAB79R	1	246
57	GNMAB80F	19923	20432
57	GNMAB80R	21103	21624
57	GNMBA07F	14530	15093
57	GNMBA07R	15847	16378
57	GNMCB11R	30694	31243
57	GNMCB47F	29518	30234
57	GNMCB47R	28242	28881
57	GNMCD55F	32780	33171
57	GNMCE88F	13260	13679
57	GNMCE88R	14546	15067
57	GNMCF06F	16859	17358
57	GNMCF06R	15242	15921
57	GNMCF40F	18554	19027
57	GNMCF40R	19698	20365
57	GNMCF50F	20435	20910
57	GNMCF50R	21576	22262
57	GNMCF63F	30402	30884
57	GNMCF63R	28818	29412
57	GNMCF86R	32361	33020
57	GNMCK71F	8763	9100
57	GNMCK71R	10055	10613
57	GNMCL95F	3811	4223
57	GNMCL95R	2299	2901
57	GNMCN67F	20529	21206
57	GNMCN67R	19529	20102
57	GNMCP09F	2860	3520
57	GNMCP09R	1894	2615
57	GNMCP70F	17618	18104
57	GNMCP70R	18924	19511
57	GNMCP79F	8875	9372
57	GNMCP79R	10275	10855
57	GNMCQ41F	20359	21104
57	GNMCQ41R	19619	20345
57	GNMCQ44F	10270	10898
57	GNMCQ44R	11575	12244

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
57	GNMCS16F	20638	20868
57	GNMCS86F	30569	31246
57	GNMCV34F	21537	21988
57	GNMCY40F	20132	20855
57	GNMCY40R	19153	19716
57	GNMCY49R	26133	26607
57	GNMCY80F	8452	8787
57	GNMCY80R	6998	7416
57	GNMCY90F	19373	19946
57	GNMCZ43F	31206	31711
57	GNMCZ43R	32436	32921
58	GNMAB82F	9525	10095
58	GNMAB82R	8509	9029
58	GNMCO58R	15112	15768
58	GNMCY78R	3411	3857
58	GNMCY83F	11793	12472
58	GNMCY83R	10643	11053
59	GNMAB85F	2737	3302
59	GNMAB85R	1900	2305
59	GNMCO33F	2304	2941
59	GNMCO33R	1257	1881
59	GNMCX86F	2826	3461
59	GNMCX86R	1441	2128
59	GNMCZ32F	1619	2126
59	GNMCZ32R	2661	3195
60	GNMAB95F	13774	14279
60	GNMAB95R	15289	15810
60	GNMCA30F	937	1556
60	GNMCD44F	303	826
60	GNMCF04F	9775	10276
60	GNMCF04R	8305	8976
60	GNMCF90F	3862	4310
60	GNMCF90R	2510	3187
60	GNMCH28F	9435	9696
60	GNMCK30F	13554	14101
60	GNMCK30R	12158	12740
60	GNMCM05F	9295	9874
60	GNMCM05R	10879	11616
60	GNMCM55F	10074	10731
60	GNMCM55R	10796	11542
60	GNMCS87F	13103	13751
60	GNMCW39F	15206	15851
60	GNMCX55F	12701	12889
60	GNMCX55R	13822	14516
60	GNMCX62R	1554	2237
61	GNMBA06F	22890	23457

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
61	GNMBA06R	24229	24758
61	GNMCB04F	30158	30722
61	GNMCB04R	28612	29214
61	GNMCB21F	23862	24428
61	GNMCB21R	25186	25806
61	GNMCB63R	3796	4094
61	GNMCB86F	23284	23998
61	GNMCB86R	24021	24623
61	GNMCD18F	31187	31608
61	GNMCF95F	20692	21018
61	GNMCF95R	19232	19872
61	GNMCK40F	11307	11811
61	GNMCK65F	9007	9517
61	GNMCL04F	20077	20543
61	GNMCL04R	18687	19271
61	GNMCL20F	27968	28464
61	GNMCL20R	29257	29840
61	GNMCL22F	13417	13939
61	GNMCL22R	14872	15438
61	GNMCL29R	34192	34771
61	GNMCL53F	1518	2034
61	GNMCL53R	214	686
61	GNMCL90R	8315	8896
61	GNMCM65F	15441	16117
61	GNMCM65R	14289	14994
61	GNMCM71F	10516	11122
61	GNMCM71R	11703	12405
61	GNMCO61F	14512	15200
61	GNMCO61R	13255	13946
61	GNMCQ79F	15902	16644
61	GNMCQ79R	16726	17426
61	GNMCQ90F	2342	3073
61	GNMCQ90R	804	1426
61	GNMCQ95F	19198	19483
61	GNMCQ95R	20653	21277
61	GNMCS24F	19718	20379
61	GNMCS46F	18786	19366
61	GNMCV12F	30913	31415
61	GNMCV13R	31908	32632
61	GNMCY25F	25038	25729
61	GNMCY25R	26701	27270
62	GNMBA12F	7833	8334
62	GNMBA66F	8661	9232
62	GNMBA66R	9606	10138
62	GNMBB30F	3235	3799
62	GNMBB30R	4483	5016

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
62	GNMCB05F	4772	5096
62	GNMCB05R	6111	6717
62	GNMCD67F	7723	8233
62	GNMCF78F	3478	3931
62	GNMCM43F	12550	13285
62	GNMCM43R	11540	12127
62	GNMCP28F	3321	3756
62	GNMCP28R	1814	2235
62	GNMCP67F	2320	2824
62	GNMCP67R	3943	4497
62	GNMCV62F	8092	8582
62	GNMCV62R	9694	10487
62	GNMCX39F	7125	7796
62	GNMCX39R	5729	6265
62	GNMCZ55F	5209	5724
62	GNMCZ55R	3782	4320
62	GNMCZ76F	4455	4947
62	GNMCZ76R	3027	3553
63	GNMBA13F	14825	15391
63	GNMBA13R	13165	13703
63	GNMBA14F	12491	13059
63	GNMBA14R	13757	14281
63	GNMBA80F	12477	12855
63	GNMCB32F	472	756
63	GNMCD42F	20565	21089
63	GNMCF07F	13708	14215
63	GNMCF07R	12522	13201
63	GNMCK47F	10432	10931
63	GNMCK47R	9275	9813
63	GNMCK91R	9054	9617
63	GNMCN32F	16696	17346
63	GNMCN32R	17927	18521
63	GNMCS55F	1461	2208
63	GNMCX85R	14727	15427
63	GNMCZ11R	17115	17610
63	GNMCZ18F	1990	2479
63	GNMCZ18R	3109	3667
63	GNMCZ34F	13696	14216
63	GNMCZ34R	12451	13003
64	GNMBA27F	2420	2987
64	GNMBA27R	649	1182
64	GNMCK68F	8858	9142
64	GNMCN47F	8600	9323
64	GNMCQ47F	5300	5761
64	GNMCQ47R	3904	4632
64	GNMCZ45F	6005	6471

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
64	GNMCZ45R	7509	8073
64	GNMCZ89F	6722	7164
65	GNMBA40F	9256	9800
65	GNMBA40R	7884	8418
65	GNMCK42F	8125	8438
65	GNMCK42R	9146	9679
65	GNMCK43F	14839	15396
65	GNMCK43R	13196	13745
65	GNMCM11R	2515	3190
65	GNMCO03F	4056	4557
65	GNMCO03R	5332	6065
65	GNMCO32F	10209	10877
65	GNMCO32R	11348	11993
65	GNMCO78R	1107	1782
65	GNMCQ10F	9012	9752
65	GNMCQ10R	10149	10831
65	GNMCQ36F	19	522
65	GNMCZ17F	1839	2369
65	GNMCZ17R	3149	3711
65	GNMCZ24R	3485	4030
65	GNMCZ50R	2017	2356
65	GNMCZ51F	3684	4187
65	GNMCZ51R	5216	5657
66	GNMBA45F	5960	6527
66	GNMBA45R	4417	4948
66	GNMBB01F	3556	4094
66	GNMBB01R	2060	2598
66	GNMCA23F	4257	4873
66	GNMCN50F	6431	7098
66	GNMCN50R	5020	5625
66	GNMCO46F	1766	2443
66	GNMCO46R	706	1195
66	GNMCQ15F	1788	2506
66	GNMCQ15R	994	1686
66	GNMCZ67F	1099	1592
66	GNMCZ67R	2554	3093
66	GNMCZ68F	1130	1584
67	GNMBA56R	828	1363
67	GNMCZ01F	1176	1497
67	GNMCZ01R	2672	3147
68	GNMBA58F	11648	12214
68	GNMBA58R	10145	10680
68	GNMBB14F	7190	7758
68	GNMBB14R	8579	9037
68	GNMCD71F	502	959
68	GNMCL54F	10328	10882

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
68	GNMCL54R	11852	12293
68	GNMCN39F	13282	13967
68	GNMCN39R	11911	12477
68	GNMCP34F	12249	12751
68	GNMCP34R	10521	11087
68	GNMCP74F	9533	10032
68	GNMCP74R	8395	8982
68	GNMCV35F	11085	11475
68	GNMCV35R	12496	12972
69	GNMBA67F	10755	11332
69	GNMBA67R	9691	10167
69	GNMCA68F	138	798
69	GNMCA95F	7720	8389
69	GNMCB19F	7635	8181
69	GNMCB62F	4968	5465
69	GNMCB62R	6482	7170
69	GNMCD88F	6048	6546
69	GNMCD94F	10463	10960
69	GNMCD94R	12298	12546
70	GNMBA87F	8256	8675
70	GNMBA87R	6890	7365
70	GNMCA76F	9130	9792
70	GNMCB96F	10306	11006
70	GNMCB96R	11786	12359
70	GNMCD20F	2427	2973
70	GNMCD20R	3980	4417
70	GNMCE77F	10510	10866
70	GNMCF49F	13718	14204
70	GNMCF49R	11782	12414
70	GNMCF57F	24615	25081
70	GNMCF57R	23522	24203
70	GNMCF81R	14890	15469
70	GNMCK10F	32790	33342
70	GNMCL64F	2279	2735
70	GNMCL64R	1098	1594
70	GNMCM94F	15929	16589
70	GNMCM94R	16990	17708
70	GNMCO70F	6253	6962
70	GNMCP46F	28269	28572
70	GNMCP46R	29399	29799
70	GNMCP69R	14839	15383
70	GNMCQ60R	4262	4932
70	GNMCV71F	1570	2085
70	GNMCV71R	316	1151
70	GNMCV72F	29887	30336
70	GNMCV72R	28290	29022

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
70	GNMCV79F	9283	9798
70	GNMCV79R	8344	9079
70	GNMCV90F	15009	15476
70	GNMCV90R	16482	17299
70	GNMCX43F	15135	15898
70	GNMCX43R	14040	14726
70	GNMCY28F	27547	28277
70	GNMCY28R	26646	27207
70	GNMCZ35F	32742	33250
71	GNMBB05F	1960	2525
71	GNMBB05R	3344	3515
71	GNMCQ43F	7860	8357
71	GNMCQ43R	8617	9224
71	GNMCV39F	3444	3908
71	GNMCV39R	1967	2637
71	GNMCV40R	1959	2698
71	GNMCX05F	7245	7867
71	GNMCX05R	9020	9558
71	GNMCY02F	11233	11831
71	GNMCY02R	10519	11074
71	GNMCZ22F	12199	12719
71	GNMCZ22R	10978	11535
71	GNMCZ62F	5934	6428
71	GNMCZ62R	7330	7740
72	GNMBB26F	8760	9327
72	GNMBB26R	7556	8099
72	GNMCA20F	13469	14085
72	GNMCA70F	3932	4596
72	GNMCA83F	16236	16703
72	GNMCD73F	16569	17077
72	GNMCD73R	15204	15432
72	GNMCF25F	16016	16451
72	GNMCF25R	14647	15269
72	GNMCM14R	10622	11346
72	GNMCS42F	5706	6424
72	GNMCS67F	9325	10026
72	GNMCS91F	3912	4620
72	GNMCY88F	1473	2157
73	GNMCA21F	82	736
73	GNMCA82F	3679	3975
73	GNMCL92F	4664	5205
73	GNMCL92R	5485	5880
73	GNMCM22R	708	1428
73	GNMCM29R	1947	2683
73	GNMCO16R	1657	2311
73	GNMCV93F	347	830

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
73	GNMCV93R	1879	2561
74	GNMCA78F	5557	6224
74	GNMCB76F	5584	6225
74	GNMCB76R	4398	4946
74	GNMCF14R	1573	2079
74	GNMCF30F	9638	10051
74	GNMCF30R	8180	8703
74	GNMCL96F	16170	16676
74	GNMCL96R	14728	15294
74	GNMCN51F	7918	8654
74	GNMCN51R	6999	7601
74	GNMCN65F	14177	14895
74	GNMCN65R	12918	13517
74	GNMCN66R	12940	13557
74	GNMCO71F	2786	3525
74	GNMCO71R	3980	4683
74	GNMCP02F	9531	10254
74	GNMCP02R	10574	11268
74	GNMCQ12F	1447	2032
74	GNMCQ12R	416	1065
74	GNMCV61F	12114	12501
74	GNMCV61R	10643	11335
74	GNMCX30F	18292	19013
74	GNMCX30R	20178	20810
74	GNMCX94F	21616	22251
74	GNMCX94R	20632	21246
74	GNMCY73R	13205	13774
74	GNMCZ16F	14762	15283
74	GNMCZ16R	13378	13933
74	GNMCZ19F	23465	23941
75	GNMCA94F	3978	4349
75	GNMCB55F	2185	2819
75	GNMCB55R	3259	3917
75	GNMCL13F	4716	5241
75	GNMCL13R	2852	3443
75	GNMCL80F	4341	4845
75	GNMCL80R	2903	3473
75	GNMCM78R	2146	2889
75	GNMCV07F	1	479
75	GNMCV08R	1221	1918
75	GNMCV10F	5011	5503
75	GNMCV11R	3483	4212
75	GNMCV36F	4495	4971
75	GNMCV36R	3285	3527
75	GNMCV52F	3868	4351
75	GNMCV52R	2491	3098

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
75	GNMCX78F	3135	3788
75	GNMCX78R	4397	5087
76	GNMCB02F	2416	2977
76	GNMCB02R	3352	3966
76	GNMCB07F	2416	2984
76	GNMCB07R	3352	3954
76	GNMCB12F	2416	2974
76	GNMCB12R	3314	3966
76	GNMCY54R	5129	5668
77	GNMCB54R	4435	4640
77	GNMCB85R	2747	3439
77	GNMCF72F	4490	4924
77	GNMCF72R	5936	6649
77	GNMCK68R	568	1128
77	GNMCM47F	3316	3922
77	GNMCM47R	4346	4995
77	GNMCX10F	6886	7627
77	GNMCX10R	5801	6436
77	GNMCZ08R	3508	3954
78	GNMCB60F	1387	2047
78	GNMCB60R	2757	3429
79	GNMCB65F	287	954
79	GNMCB65R	1598	2122
79	GNMCY11F	3301	4016
79	GNMCY11R	2339	2911
81	GNMCD15F	1	519
82	GNMCO75R	2040	2712
83	GNMCD53F	466	1013
84	GNMCF02F	1638	2132
85	GNMCF15F	3019	3523
85	GNMCF15R	1257	1932
85	GNMCY26F	1834	2612
85	GNMCY26R	555	1120
86	GNMCF34F	1890	2365
86	GNMCF34R	259	918
86	GNMCS21F	1678	2392
87	GNMCF36F	274	748
88	GNMCF71R	10636	11160
88	GNMCL78F	2657	3153
88	GNMCL78R	4106	4665
88	GNMCN10F	7355	8034
88	GNMCQ46F	10928	11579
88	GNMCQ46R	9882	10586
88	GNMCQ88F	574	1196
88	GNMCQ88R	2017	2549
89	GNMCF76F	1981	2406

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
89	GNMCF76R	1	500
89	GNMCF80F	920	1305
89	GNMCF80R	2032	2709
89	GNMCL16F	247	763
89	GNMCL16R	1226	1784
89	GNMCN80F	788	1493
89	GNMCQ82F	1969	2554
89	GNMCQ82R	401	1093
89	GNMCZ19R	2292	2850
123	GNMCH27F	119	501
145	GNMCP17R	991	1517
152	GNMCP17F	81	776
153	GNMCK10R	756	1346
153	GNMCS01F	823	1344
153	GNMCX08F	332	1001
153	GNMCX08R	1513	2144
153	GNMCZ35R	695	1204
154	GNMCK14R	1	352
155	GNMCK59R	1	445
156	GNMCK78F	8693	9133
156	GNMCM20F	2049	2694
156	GNMCM20R	632	1335
156	GNMCS15F	3468	4033
156	GNMCS66F	4788	5488
156	GNMCV01F	1890	2231
156	GNMCV02R	166	894
156	GNMCV68F	2538	3032
156	GNMCV68R	3475	4231
157	GNMCL11F	295	834
157	GNMCL11R	1294	1846
158	GNMCL30F	1756	2276
158	GNMCL30R	448	1028
158	GNMCV49R	4317	5164
159	GNMCL48F	5961	6264
159	GNMCL48R	4706	5280
159	GNMCQ61R	922	1535
159	GNMCS71F	314	1024
159	GNMCY32F	8722	9407
159	GNMCY32R	10063	10584
159	GNMCY51F	8917	9628
159	GNMCY51R	10406	10895
160	GNMCL58R	4560	5111
160	GNMCN05R	9955	10528
160	GNMCO37F	8602	9262
160	GNMCV04F	951	1370
160	GNMCV05R	1971	2742

Coordinates of Sequences Released in Contigs			
Contig No.	Sequence Name	Coordinate	Coordinate
161	GNMCN26F	4880	5549
161	GNMCN26R	3911	4533
161	GNMCQ77F	6238	6857
161	GNMCQ77R	5035	5760
161	GNMCZ58F	3859	4357
161	GNMCZ58R	2375	2916
162	GNMCN45F	1676	2346
162	GNMCN45R	400	977
163	GNMCN92F	507	1223
163	GNMCN92R	1454	2112
163	GNMCY42F	1142	1860
163	GNMCY42R	2736	3290
163	GNMCZ36F	4711	5225
163	GNMCZ36R	6070	6592
164	GNMCN94F	3000	3708
164	GNMCN94R	1705	2265
165	GNMCQ54F	51	677
165	GNMCQ54R	936	1639
166	GNMCS72F	19	432
166	GNMCS74R	1	181
167	GNMCV58F	314	808
167	GNMCZ38F	6858	7329
167	GNMCZ38R	5443	5996
168	GNMCX26F	1	660
169	GNMCX92F	341	587
170	GNMCY65R	195	567

APPENDIX B

MenB ORFs

Number 1 ORF

```
  1  ..TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
 51    GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA
101    TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
151    ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC
201    GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT
251    CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA
301    TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
351    GCCGAATAA
```

Number 2 ORF

```
  1  ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCG TCTTCTTCTT
 51    TCAGGAACGC CCCGGAAAGG ACGGAAAACC TTTTAAAATG GTCAAATTCC
101    GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
151    GAACGCCTGA CACCGTTCGG CAAAAAACTG CGTGCCGcCA GTwTGGACGA
201    ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
251    CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
301    CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG
351    GCGCAACGCg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401    TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451    AAAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACA.aCCAT
501    GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTCGGT GCGGGCGGAC
551    ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
601    ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651    CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701    ACGTCGCCGT CGCCGTCGGC AACAACCGCA TCCGCCGCCA AATCGCCGAA
751    AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
801    GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851    AAGCGGTCG..
```

Number 3 ORF

```
  1  ..AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGGT
 51    CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101    TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151    TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201    CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251    CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301    TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351    TGGCGACCCG CGTGAAGTAA GC........ .....ACCGC CGTTTCTGCA
401    CAGTTTAG
```

Number 4 ORF

```
  1  ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTC
 51  GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC
101  ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCCCCC
151  GATGCCTTCA GCGGCAATCC TGAAgGGCAG TTTTTCCCCG ACAGCTACGA
201  AATCGATGCG GGCGGCAGTG ATTTGCAGAT TTACCAAACC GCCTACAAgG
251  GCGATGCAAC GCCGCCTGAA TGAgGGCATG GGAAAGCAGG CAGGACGGGC
301  TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
351  AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401  CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
```

```
451  ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501  CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551  GATTGCGCTG CCC..
```

Number 5 ORF

```
  1  CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
 51  ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101  GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151  GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
201  GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC
251  TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC
301  GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTTG AACAGGCGGA
351  AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
401  AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
451  CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG
```

Number 6 ORF

```
  1  AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51  CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
101  CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151  TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
201  GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
251  AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
301  CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGaCTGGGCG
351  GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
401  GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT
451  TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG
501  CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GAcCGACCCG
551  ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGTT
601  CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACAACCTCC
651  TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
701  GCCCAAGGCG AAGTCGTTTC CTAA
```

Number 7 ORF

```
  1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
 51    NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101    CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151    TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201    GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251    ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301    GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351    AGGCAACCTT CTTATTATCA CACACCCTTA A
```

Number 8 ORF

```
  1  ATGTwTGATT TCGGTTTrGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
 51  GATwGtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
101  GGcTCATCGG CAGGCTGCAA CGCTTTGTCG GcAGCGTCAA ACAGGAATTT
151  GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201  AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251  TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301  CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351  TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
401  ATGCCGTC..
```

Number 9 ORF

```
  1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51  CGC.TGCGGG ACACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
101  TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201  CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
251  ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
301  CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
351  GTTTGACAGG TTTAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC
401  TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
451  TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC
501  GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
551  GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
601  CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..
```

Number 10 ORF

```
  1  ..GG.CAGCACA AAAAACAGGC GGTTGAACGG AAAAACCGTA TTTACGATGA
 51    TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT
101    ATCCCCGCGT TCGGGCTTCA AATTTTCTTC ATCCTGTTTT TAACCGCCGT
151    CGCATTCAAA ACACTGCATA CCGACCCTCA GACGGCATCC CGCCCGCTGC
201    CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTCG
251    AGCTGGGTCG GCATAGGCGG CGGTTCACTT TCCGTCCCCT TCTTAATCCA
301    CTGCGGCTTC CCCGCCCATA AAGCCATCGG CACATCATCC GGCCTTGCCT
351    GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
401    ATTGCAGGAT TGCCCGAAGG GTCACTGGGC TTCCTTTACC TGCCCGCCGT
451    CGCCGTCCTC AGCGCGGCAA CCATTGCCTT TGCCCCGCTC GGTGTCAAAA
501    CCGCCCACAA ACTTTCTTCT GCCAAACTCA AAAAATC.TT CGGCATTATG
551    TTGCTTTTGA TTGCCGGAAA AATGCTGTAC AACCTGCTTT AA
```

Number 11 ORF

```
  1  ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTGC
 51    CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCGGAATA TGTGCGTTGG
101    TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TGCGGCACTG
151    CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
201    GCTGATGGCG GTTGCCTATG TCCACCGCTG CGGTATAGAC CGGCAGCCGC
251    CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT CACGGCAGCG
301    TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
351    A
```

Number 12 ORF

```
  1  ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
 51  CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101  AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151  TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
201  CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC
251  TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
301  GGCGCGGNCG ...
```

Number 13 ORF

```
  1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
 51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
101  CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
151  CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAGCCTTTA
251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
```

351 TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTT GCTGCGGATT
401 ACGTTTCCTT ATATATTATT GATTTCCCTG TCTTCATTTG TCGGCTCGGT
451 ACTCAATTCT TATCATAAGT TCGGCATTCC GGCGTTTACG CCAC.GTTTC
501 TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
551 CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
601 ACTCGrmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
651 CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
701 GCGCCTGCgA TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
751 CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
801 ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GGCGGCACTC
851 GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
901 GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGCtgc
951 TGACGCTGCC GGCGgcGGTC GGACTGGCGG TGTTGTCGTT cCCgCtGGTG
1001 GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
1051 GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
1101 TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
1151 GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
1201 CTTTAyCGGC CCACTrrAAC rCasTCGGAC TTTCGCTTGC CATCGGTCTG
1251 GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
1301 TATTTACCAA CCTGG.CAAG GGTTGGGCAG CGTTCTT.AG CAAAAATGCT
1351 GcTCTCGCTC GCCGTGA

Number 14 ORF

1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
51 GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTCAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAArGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
401 GTCCGTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA tGGACACCAA TCCG..

Number 15 ORF

1 ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
51 GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101 TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
401 GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTTGsw CAkcATCTTT TTTGCCGCAC AGTTTGTCGC
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTTGTT TATCGGTTTT
601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
801 GrkCmmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

Number 16 ORF

1 ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
51 GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC

```
101    ATGGCATTTC GGTTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151    ATgGCTTCGC GCAGTGCGTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
201    GACGGcGgCA ATTTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTTgTT
251    CTTCAGACGG CAGCAGGTCG GTTTTGTTGT ACACCTTgAT GCACGGAaTA
301    TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351    TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
401    gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC
451    agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
501    ACT..
```

Number 17 ORF

```
  1  ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
 51    GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTTATTGTG ATGATTTTGA
101    TGCCGAACTC GGGCAGCTTC GGTTTCGGCT ATGCGTCGCT GGCGGCTTTG
151    TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC
201    GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251    AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
301    GTGGCGGCGA TTCTGCCGTT TGTGTTTGCG TATATCGGTT TGGCGAACAC
351    CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
401    GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451    GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC
501    GAATCAGGAA AAAGCCAACT GGATCGCACT CTTAAAA.CC GCGC..
```

Number 18 ORF

```
  1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGCATA CCTTGATGCT
 51  GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101  CAATCACCCG NAAACACGTT GNCAAAGACC AAATCCGNGN CTTCGGTGTG
151  GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201  CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AA.NTGACGG
251  GNATTTTGAN GGCAGGGCTG GACAAACCCT TCCAAATAGT TNAGGATACC
301  CCGAGCTATG C.TGCCACCA AGCCCTGCCG GTCAAACTCG GATCGNCTGG
351  CAGCCAGAAT...
```

Number 19 ORF

```
  1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
 51    TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
101    CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
151    GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
201    TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
251    GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
301    ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
351    AGAAAATGCC GGTGCCGCCT CTGGT..
```

Number 20 ORF

```
  1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101  ACACGCGGGC AGATGCACCG ATGCAG...
```

Number 21 ORF

```
  1  ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
 51  GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101  GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151  GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTTCTT TATTGGCTT
201  TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
```

```
251  TT..
```

Number 22 ORF

```
  1  ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
 51  TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
101  AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151  GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGCGTTCATC AGGATATTCA
201  TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..
```

Number 23 ORF

```
  1  ..TTGTTCCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
 51    GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
101    ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151    AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
201    GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
251    ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301    CTCGGTTTCC CTGTCCCCGA TGCGCGGTCG GTCATCGAAG GCCGTCTGAA
351    CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401    TCGCCTGCTA NGGCATCCTG CCGCGCCTG..
```

Number 24 ORF

```
  1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
 51    CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT.GAGTGCG
101    CCTGTTTTTC GGGTGTTTCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
151    GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
201    CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
251    CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
301    TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
351    GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTCG
401    GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
451    GTCC..
```

Number 25 ORF

```
  1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101  CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
151  CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG
201  AGAAAAAAGG CTACACCGTC AAACTGGTCG AGTTTACCGA CTATGTACGC
251  CCGAATCTGG CATTGGCTGA GGGCGAGTTG
```

Number 26 ORF

```
  1  CCTCGTCGTC CTCGGCATGC TCCAGTTTCA AGGGGCGATT TACTCCAAGG
 51  CGGTGGAACG TATGCTCGGC ACGGTCATCG GGCTGGGCGC GGGTTTGGGC
101  GTTTTATGGC TGAACCAGCA TTATTTCCAC GGCAACCTCC TCTTCTACCT
151  CACCGTCGGC ACGGCAAGCG CACTGGCCGG CTGGGCGGCG GTCGGCAAAA
201  ACGGCTACGT CCCTmTGCTG GCAGGGCTGA CGATGTGTAT GCTCATCGGC
251  GACAACGGCA GCGAATGGCT CGACAGCGGA CTCATGCGCG CCATGAACGT
301  CCTCATCGGC GyGGCCATCG CCATCGCCGC CGCCAAACTG CTGCCGCTGA
351  AATCCACACT GATGTGGCGT TTCATGCTTG CCGACAACCT GGCCGACTGC
401  AGCAAAATGA TTGCCGAAAT CAGCAACGGC AGGCGCATGA CCCGCGAACG
451  CCTCGAGGAG AACATGGCGA AAATGCGCCA AATCAACGCA CGCATGGTCA
501  AAAGCCGCAG CCATCTCGCC GCCACATCGG GCGAAAGCTG CATCAGCCCC
551  GCCATGATGG AAGCCATGCA GCACGCCCAC CGTAAAATCG TCAACACCAC
601  CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACTCAACG
```

```
651  GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....
701  ............................. GC AGACACGCCC GCCGCATCCG
751  CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC
801  ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA
851  ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA
901  TGCCCACGAA CGCCAACACC TGCGCCAAAG CCTGCTTGA
```

Number 27 ORF

```
  1  ..GAAATCAGCC TGCGGTCCGA CNACAGGCCG GTTTCCGTGN CGAAGCGGCG
 51    GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
101    GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
151    CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
201    TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
251    TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
301    GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
351    CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTCAGCCGC AACGCCTGCG
401    TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
451    GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
501    AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
551    GTTATCCTTT CCCGACCGG..
```

Number 28 ORF

```
  1  ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51  CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101  GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151  GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
201  CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251  CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301  TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351  ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401  CGGaAGAGGG CGGCGaAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451  GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501  ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551  TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601  TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC..
```

Number 29 ORF

```
   1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TCCGCmGwms TCCTGkkGTA
  51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101  GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
 151  TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
 201  CGGTTCGCtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
 251  GCCGkACTGC CCGGCGTGTT TCTGTTCGGC TTTCCCGCAC AGTTCATCAA
 301  CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
 351  GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
 401  CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
 451  GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGGC AGCGGTTTTG
 501  CCCAGCTTGC CCTGTACAAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
 551  AACCCGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
 601  AATCCaACGG GCGGGTTCGG TCAGGGATTT GGAAAGCATA GGCGGCGTAT
 651  TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
 701  GCCTTGTTTT TCCGTCAGCC GGTTCCCAAA GGCGTGGCAG AGGATGCCGT
 751  yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
 801  AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
 851  TCGATTTTTC TTGCACTGGT CATGGCACTG TATTTCGCCC GCCGTTTCGT
 901  CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
 951  ATTTCAGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
1001  ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
1051  AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
```

```
1101  GCGTGTTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
1151  TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..
```

Number 30 ORF

```
  1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101  CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151  TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201  GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
251  ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301  CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351  CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401  GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
451  CACGCGTTGG ATACG...
```

Number 31 ORF

```
  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAAyGCA GTmwrAATAT
101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201  TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251  CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301  CGCTTAGgCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
351  CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
451  TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA
501  TGGCTGCTAC GGCGTTGAT..
```

Number 32 ORF

```
  1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
 51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101  TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
151  GCTGACCGGT CTTTTATTGG CGGGCGCGGC AATGAGAAGC GGCGGGAAGG
201  TATCCGTTTA TCAGATGTTG TGGCCTATC..
```

Number 33 ORF

```
  1  ATGTTTGTTT TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
 51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
101  TCGGCAATTT GGCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151  GCG....... .....GCCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
351  ACTCGCCCGC CGCGTGCGTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
451  GATTTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA
501  ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
551  CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
601  GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
651  CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
701  ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
751  AAACACGAAG GCTTGTCCGA GTCCGCGCAA AACATCATGA AAATCCTCAC
801  AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGC
851  GCGAGGGAAA GAAAGCTTTG TACGAT..
```

Number 34 ORF

```
   1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
  51  TTTTGCGGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG ..........
                                 //
 651  .......... ...GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
 701  AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
 751  AAACCGTAA
```

Number 35 ORF

```
   1  ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC
  51  CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
 101  GCAAAAATTG CGGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
 151  TCTGTTTGCG CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
 201  CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
 251  ATGACG.... .......... .......... .......... ..........
                                 //
1201  .......... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
1251  ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301  ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
1351  GTTGTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401  CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTCAC
1451  ACGTTGGGCT ACGATATGCC GGTTTCAGGT TGTCGCGAAG GCTCGGTAAC
1501  GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
1551  AGGCGGAATA TGTTTATCCG CAATGA
```

Number 36 ORF

```
   1  ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGCG GCACACTGAC
  51     CGGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
 101     TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTTGTCCGCC
 151     yTAAAAGGAC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
 201     TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
 251     GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
 301     TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
 351     TTCCACATCG CTTTTGAACG CCCCCGCCGC CGyCyTGACG AAAAACAGCG
 401     GACGCAAAGG CGAACGcTCC GCCGGACTGT CCGTCAACGG CACGGGCGAC
 451     TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
 501     CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
 551     CCGrATACGC CGACACCGAC GTATTCGTAA CCGTCGACGT A...
```

Number 37 ORF

```
   1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
 101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
 151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
 201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
 251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
 301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
 351  ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCCT AAGCTTCTAG
 401  ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
 451  AAGATGGGTA TGCGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
 501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551  AAGTTTATGA CTTGTAysrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
 601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
 651  CGTGTTTGTC GGCCTGTCCT ATAAAATGTT GagCaGTTAC GGAAAAAAAC
 701  aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
 751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
 801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgaTTT
 851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
```

```
 901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATt
 951 gaAAGAAGTG ACGGaGTTGA TGTGccaAgG aCTATGTaAA AAacGGCTTG
1001 CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAAAG
1051 CGCGCAgCAA CATTCGGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
1101 CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GGAAACCGTT
1151 TGAAGGAATC GGaCGGGGGC GTGGTCGGAT CGGCAAACTG A
```

Number 38 ORF

```
   1 GTGGTTTTCC TGAATGCCGA CAACGGGATA TTGGTTCAGG ACTTGCCTTT
  51 TGAAGTCAAA CTGAAAAAAT TCCATATCGA TTTTTACAAT ACGGGTATGC
 101 CGCGTGATTT CGCCAGCGAT ATTGAAGTGA CGGACAAGGC AACCGGTGAG
 151 AAACTCGAGC GCACCATCCG CGTGAACCAT CCTTTGACCT TGCACGGCAT
 201 CACGATTTAT CAGGCGAGTT TTGCCGACGG CGGTTCGGAT TTGACATTCA
 251 AGGCGTGGAA TTTGGGTGAT GCTTCGCGCG AGCCTGTCGT GTTGAAGGCA
 301 ACATCCATAC ACCAGTTTCC GTTGGAAATT GGCAAACACA AATATCGTCT
 351 TGAGTTCGAT CAGTTCACTT CTATGAATGT GGAGGACATG AGCGAGGGCG
 401 CGGAACGGGA AAAAAGCCTG AAATCCACGC TGCCCGATGT CCGCGCCGTT
 451 ACTCAGGAAG GTCACAAATA CACCAAT... .......... .....TACCG
 501 TATCCGTGAT GCGCCAGGCC AGGCGGTCGA ATATAAAAAC TATATGCTGC
 551 CGGTTTTGCA GGAACAGGAT TATTTTTGGA TTACCGGCAC GCGCAGCGC.
 601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
 651 AGCGGACACC TTTATGGCAT TGCGTGAGTT TTTGAAAGAT GGGGAAGGGC
 701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
 751 GAACAATTCA TGCTGGCTGC GGAAAACACG CTGAACATCT TTGCACAAAA
 801 AGGCTATTTG GGATTGGACG AATTTATTAC GTCCAATATC CCGAAAGAGC
 851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
 901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
 951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
1101 TTTGGTCTAT CTC...
```

Number 39 ORF

```
   1 ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT
  51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
 101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
 151 GyCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
 201 CGATAATCAG ACCATCGAGA ACAAACTGGA AATATTTGTC TCAGGCTATA
 251 AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA
```

Number 40 ORF

```
   1 ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
  51 CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
 101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
 151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
 201 GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
 251 AACAAGCGTT GGCCn.AGAA TTTCAACCC...
```

Number 41 ORF

```
   1 ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
  51 CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
 101 TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGGnnn nnnnnnnnnn
 151 nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
```

```
 201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
 251  AAGTCATCGT CTTCGGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
 301  GACCCCGCCT TCGCCCTGCA ACTGCCCCTA CGCGTCCTCG TTACCGAAAC
 351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
 401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
 451  AAACTGATAC AAAAAACCGT AGGCGAATAA
```

Number 42 ORF

```
   1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
  51  GCTGATGCTC CCCGCCTTTC AGAATGTGGC GGCGGAGGGG ATAGATGTGA
 101  GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
 151  CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGg
 201  CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
 251  CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
 301  GACTACAAAC TGAGTTTCCA TCCGCTGACc AaACGCTACC GCGTTACCgT
 351  CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
 401  CCGGCGCGGT TGCCAACTGG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
 451  GCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC
 501  AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACTGGC
 551  ATTTGGATTC GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA
```

Number 43 ORF

```
   1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
  51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
 101  ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTgACGGTG
 151  TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
 201  CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTTCCTG CCGCCGCTGC
 251  TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
 301  TCTGAAATCC TGTTTTCACT CGACGATGCC gCCGCCGGCa TCGGGCTGGT
 351  GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTGC
 401  GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCcCAAG
 451  CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
 501  AGCGAACACC GCCGTCCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
 551  CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
 601  CACGCACCGT TTTCGCCCGC CGTCCTGCAC CGGGGG.TGC GCTACGGCAT
 651  ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
 701  GTTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCG
 751  ATGGGTATTT CGTTCGGCGG GGCGGCATTA TTGTTCCAAA GCATCTTTTC
 801  AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
 851  CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCTCC
 901  GCCCTCTGC. TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
 951  GGAAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATG.TGCCGC
1001  CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
1051  CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101  CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG
1151  CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
1201  GAAAGCTCyT GCCGCCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT
1251  GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301  CGCCGGCAAA CTATCCCCTG TTTGCCGGCG TATGGGCGGC ATATCTGGCA
1351  GGCTGCATCC TGCGCCACCG GAAAGATTTG CACAAACTGT TTCATTATTT
1401  GAAAAAACAA GGTTTCCCAT TATGA
```

Number 44 ORF

```
   1..ATCCTGAAAC CGCATAACCA GCTTAAGGAA GACATCCAAC CTGATCCGGC
  51  CGATCAAAAC GCCTTGTCCG AACCGGATGC TGCGACAGAG GCAGAGCAGT
 101  CGGATGCGGA AAATGCTGCC GACAAGCAGC CCGTTGCCGA TAAAGCCGAC
 151  GAGGTTGAAG AAAAGGCGGG CGAGCCGGAA CGGGAAGAGC CGGACGGACA
 201  GGCAGTGCGT AAGAAAGCGC TGACGGAAGA GCGTGAACAA ACCGTCAGGG
 251  AAAAAGCGCA GAAGAAAGAT GCCGAAACGG TTAAAATACA AGCGGTAAAA
```

```
301  CCGTCTAAAG AAACAGAGAA AAAAGCTTCA AAAGAAGAGA AAAAGGCGGC
351  GAAGGAAAAA GTTGCACCCA AACCAACCCC GGAACAAATC CTCAACAGCG
401  GCAgCATCGA AAAmGCGCGC AgTGCCGCCG CCAAAGAAGT GCAGAAAATG
451  AA.AACGTCC GACAAGGCGG AAGC.AACGC ATTATCTGCA AATGGGCGCG
501  TATGCCGACC GTCAGAGCGC GGAAGGGCAG CGTGCCAAAC TGGCAATCTT
551  GGGCATATCT TCCAAGGTGG TCGGTTATCA GGCGGGACAT AAAACGCTTT
601  ACCGGGTGCA AAGCGGCAAT ATGTCTGCCG ATGCGGTGA
```

Number 45 ORF

```
  1  ATGAACCACG ACATCACTTT CCTCACCCTG TTCCTACTCG GTkTCTTCGG
 51  CGGAAcGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GcGTTTGs.s
101  TCCAACTCCC CCCGCATATC AACCGCTTTT GGCTGATCCT GCTGCTTAAC
151  ACAGGACGGG TAAGCAGCTA TACGGCAAtC GGCCTGATAC TCGGATTAAT
201  CGGACAGGTC GGCGTTTCAC TCGAcCAaAC CCGCGTCCTG CAGAATATTT
251  TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301  GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAaATCGGCA AACCGATATG
351  GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
401  CCGCCTGCCT tGCGgTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451  GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AgCGGTAGTG CGGCAACGGG
501  CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTtTAG
551  CAATCGGCAT TTTtTCCCTG CAACTGAAwA AAATCATGCA AAACCGATAT
601  ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
651  TGCCGTCCTG TGGCTGTAA
```

Number 46 ORF

```
  1  ATGGAAAACC AAAGGCCGCT CCTAGGCTTT CGCTTGGCAC TTTTGGCGGC
 51  GATGACGTGG GGAACGCTGC CGAT.TCCGT GCGGCAGGTA TTGAAGTTTG
101  TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
151  TTGTTTGTTT TGCTGGCACT GGGCGGGCGG CTGCcGAAGC GGCGaGGATT
201  TTTCTTGGTG CTCATTCAGG CTGCTGCTGC TCGGCGTGGC GGGCATTTCG
251  GCAAACTTTG TGCTGATTGC CCAAGGGCTG CATTATATTT CGCCGACCAC
301  GACGCAGGTT TTGTGGCAGA TTTCGCCGTT TACGATGATT GTwGTCGGTG
351  TGTTGGTGTT TAAAGACCGG ATGACTGCCG CTCAGAAAAT CGGCTTGGTT
401  TTGCTGCTTG CCGGTTTGCT TATGTATTTT AACGATAAAT TCGGCGAGTT
451  GTCGGGTTTG GGCGCGTATG C.AAGGGCGT GTTGCTGTGT GCGGCAGGCA
501  GTATGGCATG GGTGTGTAAT GCCGTGGCGC AAAAGCTGCT GTCGGCGCAA
551  TTCGGGCCGC AACAGATTCT GCTGTTGATT TATGCGGCAA GTGCCGCCGT
601  GTTCCTGCCG TTTGCCGAAC CGGCACACAT CGGAAGTATG GACGGTACGT
651  TGGCGTGGGT ATGTATTGCG TATTGCTGCT TGAATACGTT AATCGGTTAC
701  GGCTCGTTCG GCGAGGCGTT GAAACATTGG GAGGCTTCCA AAGTCAGCGC
751  GGTAACAACC TTGCTCCCCG TGTTTACCGT AATAAATACT TTGCTCGGGC
801  ATTATGTGAT GCCTGAAACT TTTGCCGCGC CGGA..
```

Number 47 ORF

```
  1  ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
 51  .CCCGCAACC TAATTTCAAA CCCCTCGGTT CAATGCCGAG GG.GTTTTGT
101  T.TTGCCTGT TTCCTGTTTC CTGTTTCCTG CCGCCTCCGT TTTTTGCCGG
151  ATTTTCCTTC CGGCCGCAAT ATCGGAACGG CAGACCGCCG TCTGTTTGCG
201  GTTGCAAATT CAGGCAGTTT GGCTACAATC TTCCGCATTG TCTTCAAGAA
251  AGCCAACCAT GCCGACCGTC CGTTTTACCG AATCCGTCAG CAAACAAGAC
301  CTTGATGCTC TGTTCGAGTG GGCAAAAGCA AGTTACGGTG CAGAAAGTTG
351  CTGGAAAACG CTGTATCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
401  AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGGCTG CyCGGAGTCT
451  TCAGACGGCA TTTTTCTGAA TgCGGACGGc TGgCctGATA TGGgCGGAcg
501  cTTACAGCAC CTCGCCCTCG GTTGGCACTG TGCGGGGCTG TTGGACGgsT
551  GGCGCAACGA GTGTTTCGAC CTGACCGACG GCGGCGGCAA CCCCTTGTTC
601  ACGCTCGaAc GCGCCGyTTT mCGTCCTkTC GGACTGCTCA GCCGCGCCGT
651  CCATCTCAAC GGTCTGACCG AATCGGACGG CCGATGGCAT TTCTGGATAG
701  GCAGGCGCAG TCCGCACAAA GCAGTCGATC CCAACAAACT CGACAATACT
```

751 rCCGCCGGCG GTGTTTCCGG CGGCGAAATG CCGTCTGAAG CCGTGTGTCG
801 CGAAAGCAGC GAAGAAGCCG GTTTGGATAA AACGCTGcTT CCGCTCATCC
851 GCCCGGTATC GCAGCTGCAC AGCCTGCGCT CCGTCAGCCG GGGTGTACAC
901 AATGAAATCC TGTATGTATT CGATGCCGTC CTGCCG...

Number 48 ORF

1 ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51 CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
151 TTGATATTTG GTAACTATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
201 ACCTGCATCG GGCGTAATCA GGGTGTATGC ACCGGgATACG rGkACAATTA
251 CAGCGAAATT CGTGGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAACTGG
401 GTCGTCTGAA GCTGATACAC GGGAATGAAA CGCGCAgCcT TAAAGCAACT
451 GTCGAACGTT TGGAAAACCA GGAACTCCAT ATTTCGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

Number 49 ORF

1 ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
51 GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAAACCG ATGATAAGGA CAGTCCGGCA GGTTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

Number 50 ORF

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
51 CGgATTTATC GATgcgatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAgCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCAC TGTATTTTGT
351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTTCTGTT cGGGCTGACG GTCGC.ACCG CTTTTGGGTT TTTACGACGG
451 TGTGTTCGGA CCGGGTGTCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
501 TGCTCGGCTG CAAgCTGTTG AACGCGATGT CTTACACCAA ATTGGCGAAC
551 GTTGCCTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTCGAT
601 TATTTTCCCG ATTGCGGCAA CGaTGGCGGT CGGTGCGTTT GTCGGtGCGA
651 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

Number 51 ORF

1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
51 CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
101 TGGTTTTCTG GGACTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
151 TCATGATGTT TTTGGTGGTT TCTACCAGTT TGTGCCTGAT TCGCAATGTG
201 CCGCCGTTCT GGCGCGAAAT GAAGTCTTTT CGGGAAAAGG TTAAAGAAAA
251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGGATGTA AAAATTGCGC
301 CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTCA GGGGAAAACC
351 ATTAACCGTG AAGACGGGTC GGTTCTGATT GCCGCCAAAA AAGGCACAAT

```
401   GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC
451   TGGGCGGGTT GATAGACAGT AACCTGCTGT TGAAACTGGG TATGCTGACC
501   GGTCGGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTTC.AAGC
551   CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA
601   TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA
```

Number 52 ORF

```
   1  ATGCCGTCTG AAACACGCCT GCCGAACTTT ATCCGCGTCT TGATATTTGC
  51  CCTGGGTTTC ATCTTCCTGA ACGCCTGTTC GGAACAAACC GCGCAAACCG
 101  TTACCCTGCA AGGCGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
 151  TCAAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAAACGCAT
 201  CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
 251  ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCTCCGC
 301  ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC GCCTGAACCG
 351  CCTGACACAC GGCGCGCTGG ACGTAACCGT CGGCCCCTTG GTCAACCTTT
 401  GGGGATTCGG CCCCGACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA
 451  ATCAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGAAACA
 501  AGGCAAAGAT TACGCTTCCT TGAGCAAAAC CCACCCCAAG GCCTATTTGG
 551  ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TGCGGGCGAA
 601  CTGGAAAAAT ACGGCATTCA AAATTATCTG GTCGAAATCG GCGGCGAGTT
 651  GCACGGCAAA GGCAAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
 701  AGCAGCCCAA TATCGTCCAA GGCGGCAATA CGCAGATTAT CGTCCCGCTG
 751  AACAACCGTT CGCTTGCCAC TTCCGGCGAT TACCGTATTT TCCACGTCGA
 801  TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
 851  CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGCAGA CAGTGCGATG
 901  ACGGCGGACG GCTTGTCCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
 951  CTTAAAGCTG GCAGAGCGCG AAAAACTCGC TGTTTTCCTG ATTGTCAGGG
1001  ATAAAGGCGG CTACCGCACC GCCATGTCTT CCGAATTTGA AAAACTGCTC
1051  CGCTAA
```

Number 53 ORF

```
   1  ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCGTCA
  51    AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
 101    GCGGCGCGGC TGCGGACGGG TGGCGCAAAG GCGTGCAAAT CGGCGGCGAG
 151    GTGTTTGTAC GGCAAAATGA AGGCAGCCkA yTGGCAATCG GCGTGATGGG
 201    CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
 251    gCAGTGATTT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
 301    TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
 351    ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
 401    CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
 451    GAAGGCATTG TCGGAAAAGG CAATAATGTG CGGTTTTACC TACAACCGCA
 501    GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTTACC GACAGCGAGG
 551    GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
 601    AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTCA ATCTTCAGCC
 651    TTTTGCCGCT TTTAATGTTt TGCACAGGTC AAAATCTTTC GGCGTGGAAA
 701    TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG
 751    TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..
```

Number 54 ORF

```
   1  ..GCGGAATATG TTCAGTTCTC TATAGATTTG TTCAGTGTGG GTAAATCGGG
  51    GGGCGGTATA CCTAAGGCTA AGCCTGTGTT TGATGCGAAA CCGAGATGGG
 101    AGGTTGATAG GAAGCTTAAT AAATTGACAA CTCGTGAGCA GGTGGAGAAA
 151    AATGTTCAGG AAACGAGAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
 201    TGCGCAACGA GAATGGGAAA ATAAAACAGG GTTAGATTTT AATCATTTTA
 251    TAGGTGGTGA TATCAATAAA AAAGGCACAG TAACAGGAGG GCATAGTCTA
 301    ACCCGTGGTG ATGTACGGGT GATACAACAA ACCTCGGCAC CTGATAAACA
 351    TGGGGT.TTA TCAAGCGACA GTGGAAATTN A
```

Number 55 ORF

```
   1  ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCT
  51  GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101  TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151  TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
 201  CAAAATTGCC GGCGTATTGG CGTTTTGGCT GGCGGTTTTG TTTGACGGGC
 251  TGATGATGGT GATCCAACTC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301  AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
 351  CGGGCTG...
```

Number 56 ORF

```
   1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
  51    TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
 101    GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
 151    GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
 201    AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
 251    CTTCCGAATA CCGCGACGGG ATTTTTGAAT TCAACGTCGG TTATATCGCC
 301    AGTGCGGTTT TGGCTTTGGT TTTCCTTGCA CTGGGCGC.G TAGCGCCGAA
 351    CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAAATAT AACGGGCAAT
 401    TGATCAATAT GTACGCC..
```

Number 57 ORF

```
   1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTTGT
  51    TGCGCTTGCC GGCTTGTTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGA
 101    TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
 151    CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
 201    GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTCG
 251    TCCGTTTCTG CCGAAACTAT CTGGCGCACG AATCCGAACC GGACAGGCCC
 301    GTTCCGCCT..
```

Number 58 ORF

```
   1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
  51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
 101  TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
 151  GCATTGGTCG GCTTCTGGGT C......... .......... ..........
                                    //
 901  .........A TTGCCATCGG TTTGTTTTTA ATTTACCAAA ACGGGCTGAC
 951  CCTGCTTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
1001  TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
1051  GTCCGCAGTA TGCCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
1101  GACATTGAAA GGCGGAAAAT GA
```

Number 59 ORF

```
   1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
  51    GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
 101    TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
 151    ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
 201    TCGTGTCGTC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
 251    CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
 301    AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
 351    TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
 401    ATTCGTAA
```

Number 60 ORF

```
  1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
 51    CAACATTTCA CTGGGTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
101    GCCATCATGC GCCCAGCCAA GGCACTGAGT TGCCGCAAAG CAACGGTATT
151    TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201    ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
251    GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCtGGACAGC
301    CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
351    CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
401    GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
451    AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
501    AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
551    AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
601    CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
651    GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
701    CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
751    GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
801    ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
851    TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901    CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
951    TATCACAGGC AAAGAAAAAG GTGTTT..
```

Number 61 ORF

```
  1  ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
 51    TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
101    ACACGACCCA TGTTGATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
151    GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
201    CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
251    ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
301    CAAGCAGGCA ATCATGTTCG CATTGGTACA ACCCAAACTC AAAGCCAAAG
351    CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
401    GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
451    AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
501    TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
551    AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
601    AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
651    GGTGGCATTC AGTTCGCCCG TTACCGATTT GGCACAACAA ...
```

Number 62 ORF

```
  1  ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51  CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
101  GACACTCCGA CAAAGATGCC CTGCTCAACA GCAwAACCAG CCATGTCCGC
151  GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
201  GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGyCATGCGC AACCTGCAAG
251  AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301  TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351  CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
401  CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA
451  CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
501  CGTGCGCATC GACTTCATCT CCTAT...
```

Number 63 ORF

```
  1  ..GCGCGGCACG GCACGGAAGA TTTCTTCATG AACAACAGCG ACAC.ATCAG
 51    GCAGATAGTC GAAAGCACCA CCGGTACGAT GAAGCTGCTG ATTTCCTCCA
101    TCGCCCTGAT TTCATTGGTA GTCGGCGGCA TCGGCGTGAT GAACATCATG
151    CTGGTGTCCG TTACCGAGCG CACCAAAGAA ATCGGCATAC GGATGGCAAT
201    CGGCGCGCGG CGCGGCAATA TTTyGCAGCA GTTTTTGATT GAGGCGGTGT
251    TAATCTGCGT CATCGGCGGT TTGGTCGGCG TGGGTTTGTC CGCCGCCGTC
```

301 AGCCTCGTGT TCAATCATTT TGTAACCGAC TTCCCGATGG ACATTTCCGC
351 CATGTCCGTC ATCGGCGCGG TCGCCTGTTC GACCGGAATC GGCATCGCGT
401 TCGGCTTTAT GCCTGCCAAT AAAGCAGCCA AACTCAATCC GATAGACGCA
451 TTGGCACAGG ATTGA

Number 64 ORF

1 ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
51 GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCTTCCT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGCC GGCGCGATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

Number 65 ORF

1 ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCCGCAGA TAATCCGAGT
51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
151 CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TGCGGCATCG TGTTCGGTGC
201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAGTTCG CCsGGTTCAT
351 TGTTCAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
401 CACATATGTT CGCAAATTTC GCCGTCTTCG CCGTCTTGGA AAAAAGGGAC
451 TTTGACCATG GCAAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCCAAA
501 AAAGcTCGCG CCAAAAAATAT TTGAATGTTT TACGGGCGCG TTCGTCGGCA
551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG
651 GCTTTCTgcC kTCGGCATCC GATTCGGATT TGAAAAGTTC mmrwyATTCG
701 GAATAG

Number 66 ORF

1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
51 CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GGCGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
251 CGATTGTCGG CAACCTTTTT GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAAATTAC ATCAACCGAA
401 AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC..

Number 67 ORF

1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGcTG CCGCTTTCCT
101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA
201 CCCCGACCCC AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG
251 GTTTGGAACT TGCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
351 ACACGAAGGG CTGCTATTC..

Number 68 ORF

```
  1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
 51    GCATGCGGTG TGGAATACTT TGCGCTTCTC GGCGGCGGCG GTGTATGCGG
101    CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CGGCGCGGCG GTCGGCGTGG
151    ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
201    GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
251    TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
301    TTATCAGCCT GGGATGCACT GCCGCCGGAT TACGGCAGGG CGGCGGCGGG
351    TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
401    TGAAACCGGC GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
451    GGCGAATTTG CGGCGACATT GTTTCTGTCG CGTCCGGAAT GGCAGACGCT
501    GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
551    CGCGGGCGAT GGTGCTG..
```

Number 69 ORF

```
  1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
 51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
101  ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
151  GGTTTGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
201  CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
251  AACGTTTGGT C...
```

Number 70 ORF

```
  1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
 51    TTTGCTGTCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
101    GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
151    AACTTTTTGG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
201    CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
251    CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
301    CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
351    GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
401    TACTGATGTT TTTCCGTCCG ..
```

Number 71 ORF

```
  1  ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
 51    GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101    CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151    AGCGGTTTTC AGGTAGGCTA TACGTTTTAA
```

Number 72 ORF

```
  1  ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC
 51  GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAACTCGA
101  ATTCCTCTAT GGCGATGAAA ACGGTCATTC AGACGGCATC AATTTGwCGG
151  ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCCGGCAG CGGTAAGGCG
201  TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
251  TGAGGCGGCG GAAGAGTTGG GGTTGTTGGC GGCAGAAGTC GCACAGATGG
301  AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC..
```

Number 73 ORF

```
  1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGr
```

```
101  CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCGTGCTG
151  ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGCTGGTC
201  GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
251  ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301  ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
351  GACGATAGAC AATACGTTCA ACCGCATCTG GaCGGGTCAA wTyCCAGCGT
401  CCGTGGATG..
```

Number 74 ORF

```
  1  ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 51    AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101    GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
151    ACCCGCCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG
201    CCTGCTTGAA ACACGGGAAC ACGGCTGA
```

Number 75 ORF

```
  1  ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
 51    GGGCAAACTC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA
101    AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTTCCGAT
151    GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT
201    GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
251    TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
301    GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG
351    GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG
401    GTGCAGCGCT TGCCGATATG GCGGAACTGT TCCCCGAACG CCGATTAATG
451    CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT
501    TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
551    AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG
601    TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC
651    CAAACAGGCG GCGGAGCTTG CTGCCAAAAT CACGGGCGAG GGAAAGAAAG
701    CTTTGTACGA T..
```

Number 76 ORF

```
   1  ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
  51  AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
 101  TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
 151  TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
 201  GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
 251  CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
 301  GTGGCGGcAT TGGTGGGCGt ATCAATATAT TGTGAGCGTG GCACATAACG
 351  GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAAk.AA tATCCC.GAT
 401  CAACAwCGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
 451  GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCATA
 501  AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
 551  CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTC GTATTGGGGC
 601  AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA
 651  GTTCATATCA TATTGCAAGT .......... .......... ..........
 701  .......... .....GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
 751  AAAGTGGTTA ATTAATGGGG TATTGCAAAC GGGCAACCCC TATATAGGAA
 801  AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
 851  TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACGTC AAAATGGGAA
 901  ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC
 951  ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG
1001  TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
1051  AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT
1101  CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT
1151  CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA
1201  AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
1251  CCGTTACTTG GAAAGTAAAC GGCGTGGCAA ACGACCGCCT GTCCAAAATC
```

```
1301  GGCAAAGGCA CGCTG..... .......... .......... ..........
                           //
2101  .......... .......... .......... .......... ...GATAAAG
2151  TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
2201  GATCACGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT
2251  TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
2301  ACGGCAACCk TAgCCtCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC
2351  ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG
2401  CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCCGGC AACGCTAAGG
2451  CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG
2501  GCAGTATTCC ATTTTGAAAG CAGCCGCTTT ACCGGACAAA TCAGCGGCGG
2551  CAagGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCAg
2601  GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT
2651  TCCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA
2701  TGCGCCGCGC CGCCGTTCGC GCCGTTCGCG CCGTTCCCTA TTATmCGTTA
2751  CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC
2801  AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA
2851  CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT
2901  TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG
2951  GTAGTGGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTCAC
3001  CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG...... ..........
                           //
3551  .......... .......... ....TTAGAC CGCGTATTTG CCGAAGACCG
3601  CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAAA CACTACCGTT
3651  CGCAAGATTT CCGCGCCTAC CGCCAACAAA CCGACCTGCG CCAAATCGGT
3701  ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTCGCACAA
3751  CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG
3801  CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC
3851  ATCAGnCGCG GGCGCGGGTT TTAGCAGCGG CAGCCTTTcA GACGGCATCG
3901  GAGsmAAAwT CCGCCGCCGC GTGCtGCATT ACGGCATTCA GGCACGAtAC
3951  CGCGCCGGtt tCggCGgATt CGGCATCGAA CCGCACATCG GCGCAACGCg
4001  ctATTTCGTC CAAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
4051  CCCCCGGCCT TGCATTCAAC CGcTACCGCG CGGGCATTAa GGCAGATTAT
4101  TCATTCAAAC CGGCGCAACA CATTTCCATC ACGCCTTATT TGAGCCTGTC
4151  CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG
4201  TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC
4251  GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG
4301  CCCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT
4351  GGTAA...
```

Number 77 ORF

```
  1  ..AAGGTGTGGC AATTTGTCGA AGA.CCGCTG CGTGCCGTCG TGCCTGCCGA
 51    CAGTTTTGAA CCGACCGCGC AAAAATTGAA CCTGTTTAAG GCGGGTGCGG
101    CAACCATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
151    CAGTTCCCTG CTTATGCCGC TAACTTCCCC GTTTGGGCGg ATCAGGCAAA
201    CGCGATGGTG CAGTATGCCG TTTGGACGAC ACTTGCCGCG GTCGGCGTAG
251    GTGCAAACCT GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
301    GCGTGGAATA TCCCCGAAAA CTGGTTGTTG CGCGCACAAA TGGTTATCGG
351    CGGTATTGAA GGGGCGGCAG GTGAAAAGAC CTTTGAACCC GTTGCAGAAC
401    GTTTGAAAGT GTTCGGCGCA TAA
```

Number 78 ORF

```
  1  ..GGCTACAACT ACCTGTTCGC GCGCGGCAGC CGCATCGCCA ACTACCAAAT
 51    CAACGGCATC CCCGTTGCCG ACGCGCTGGC CGATACGGGt CAATGCCAAC
101    ACCGCCGCCT ATGAGCGCGT AGAAGTCGTG CGCGGCGTGG CGGGGCTGCT
151    GGACGGCACG GGCGAGCCTT CCGCCACCGT CAATCTGGTG CGCAAACGCC
201    TGACCCGCAA GCCATTGTTT GAAGTCCGCG CCGAAGCgGG CAACCGcAAA
251    CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC
301    rCTGCGCgGC CGCCTGGTTT CCAcCTTCGG ACGCGGCGAC TCGTGGCCGC
351    GGCGCGAACG CAGCCGskAT GCCGAACTCT ACGGCATTTT GGAATACGAC
401    ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCCAA
```

```
451    AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT
501    ATGCCACCGC CTTCGGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC
551    AGCCACCACC GTGCGCTCAA CCTGTTCGCC GGCATCGAAC ACCGCTTCAA
601    CCAAGACTGG AAACTCAAAG CCGAATACGA CTAC..
```

Number 79 ORF

```
  1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTCGCCG GGAACGGCAA
101  TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
151  AGCGTCAgcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA
201  AACGGGGATA AACGcGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA
251  TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301  CCGTGCGTAC CGCAGACGCT CAAGCCCATT TnTTCAAGAA TGCGTGCCAC
351  TnAGTCGCCG ACGGGG..
```

Number 80 ORF

```
  1  ..ACCGACGTGC AAAAAGAGTT GGTCGGCGAA CAACGCAAGT GGGCGCAGGA
 51    AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
101    ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151    ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG
```

Number 81 ORF

```
   1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
  51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
 101  GCATCGGTAT TCTGGwysGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
 151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
 201  CGsyGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CkGATACTTT
 251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.........
                                //
 851  .......... .......... .......... ........AC TTCGCTGGTA
 901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
 951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
1001  TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451  TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTTCT GTTGAAAGAT
1501  AAAAAA..
```

Number 82 ORF

```
  1  ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51    CGATGAGCCT GCCAAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT
101    CGGAACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
151    TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
201    GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCCTGA
```

Number 83 ORF

```
  1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
 51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
```

```
101  A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151  TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201  GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251  GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401  TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501  AGACCG...
```

Number 84 ORF

```
  1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
 51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101  GCTACACGGG AACGCACkAG CTGTCCGGTT TCTATTGGCA CGCGCATGAg
151  ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTaTCTGGTC
251  GGCTTGACTA TCTTTTGGCT GGCTGCGCGG ATTGCCGCCT TTATCCCGGG
301  TTGGGGTGCG TCGGCAAGCG GCATACTCGG TACGCTGTTT TTCTGGTACG
351  GCGCGGTGTG CATGGCTTTG CCCGTTATCC GTTCGCAGAA TCAACGCAAC
401  TATGTTgCCG TGTTCGCGCT GTTCGTCTTG GGCGGCACGC ATGCGGCGTT
451  CCACGTCCAG CTGCACAACG GCAACCTAGG CGGACTCTTG AGCGGATTGC
501  AGTCGGGCTT GGTGATG
```

Number 85 ORF

```
  1  ..ATGCCGTCTG AAGGTTCAGA CGGCmTCGGT GyCGGGGAAy CAGAAGyGGT
 51    AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
101    CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
151    GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
201    TGTkGCTTTC GTGATAGGsA GGTTTGyTGG kmksAsyTTG TAyrATwkkG
251    CCTssCwsTG kAGmGCCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTTTy
301    TkTTyyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
351    CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
401    GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
451    TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTTTAG
501    TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
551    ATTCTCCAGC CGCCGAAATC ..
```

Number 86 ORF

```
  1  ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
 51  TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101  TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151  CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201  CATGTTCGCC GCCGGACGAA TTTGGGGGCA GArArTCCTA rGGTTCArAC
251  CTATTGCGsG CATCATGACG CCGrAACGTT ATGAGCAGGT TCAGGAAAAA
301  TTCGACAAAT ACGGTAACTG GGTCTTATTT GTCGCCCGTT TCCTGCCCGG
351  TTTGAGAACG GCCGTATTTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
401  ACTTGCGTTT TATCATTATG GATGGACTGG CCGCA...
```

Number 87 ORF

```
  1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
 51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101  AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151  AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201  AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251  AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301  AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351  TAAAATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
```

401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...

Number 88 ORF

1 ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
51 ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151 CCGAAGCAAT GGCGGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
251 CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
301 CGGATTCCGG TTGTGAAAtC CATCTATTCG AGTGTGAAAA AAGTATCCGA
351 ATacgTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GTACTCGTGC
401 CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
451 TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTTCCGT
501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
601 TATGTGATTT CGCTGGGTAT GGTCATCCCT GACGACCTGC CCGTCAAAAC
651 ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
701 AA

Number 89 ORF

1 ATgAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT
51 GGCGCTGGCT TCGGGCATTT ACACCGGCGA CGTGTATATC GTACTCGGAC
101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GsGgTACTCA
201 ATATCCCCGA AAAGATGCAG CGTTTCGGTT CGGCnCGTAA AGGCCkCAAG
251 ssCGsGCTTG CCTTGAACAA GGCGGGTTTG GCGTATTTTG AAGGGCGTTT
301 TGAAAAGGCG GAACTAGAAG CCTCACGCGT GTTGGTCAAC AAAGtAGGCC
351 GaGAGACAAC CGGACTTTGG CATTGATGCT GrGCGCGCAC GCCGCCGGAC
401 AGATGGAAAA CATCGAssTG CGCGACCGTT ATCTTGCGGA AATCGCCAAA
451 CTGCCGGAAA AACAGCAGCT TTCCCGTTAT CTTTTGTTGG CGGAATCGGC
501 GTTGAACCGG CGCGATTACG AAGCGGCGGA AGCCAATCTT CATGCGGCGG
551 CGAAGATGAA TGCCAACCTT ACGCGCCTCG TGCGTCTGCA .ATTCGTTAC
601 GCTTTCGACA GGGGCGACGC GTTGCAGGTT CTGGCAAAAA CCGAAAAACT
651 TTCCAAGGCG GGCGCGTTGG GCAAATCGGA AATGGAACGG TATCAAAATT
701 GGGCATATCC GTCGCCAGCT GGCGGATGCT GCCGATGCCG CCGCTTTGAA
751 AACCTGCCTG AAGCGGATTC CCGACAGCCT CAAAAACGGG GAATTGAGCG
801 TATCGGTTGC GGAAAAGTAC GAACGTTTGG GACTGTATGC CGATGCGGTC
851 AAATGGGTCA AACAGCATTA TCCGCAsAAC CGCCGCCCCG AGCTTTTGGA
901 AGCCTTTGTC GAAAGCGTGC GCTTTTTGGG CGAGCGCGAA CAGCAGAAAG
951 CCATCGATTT TGCCGATGCT TGGCTGAAAG AACAGCCCGA TAACGCGCTT
1001 CTGCTGATGT ATCTCGGTCG GCTCGCCTTC GGCCGCAAAC TTTGGGGCAA
1051 GGCAAAAGGC TACCTTGAAG CGAGCATTGC ATTAAAGCCG AGTATTTCCG
1101 CGCGTTTGGT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAGAAG
1151 GCGGAGGCGC AC...

Number 90 ORF

1 ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
51 GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
201 CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
251 ACGTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
301 GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
401 TGTATsTGGT CGTGTTCAAA CCGTTTTGA

Number 91 ORF

```
  1  ATGGCAAAAA TGATGAAATG GGCGGCTGTT GCGGCGGTCG CGGCGGCAGC
 51  GGTTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
101  TTACGGAAAC GGTCAGGCGC GGC // .....
//.. ATTTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
 51  CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
101  GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
151  GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
201  GGTTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
251  TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTTGGG TGCGGACGGC
301  AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
351  CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
401  TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCG
451  CCGCGCCGAT AA
```

Number 92 ORF

```
  1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAAATCGT
 51    TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
101    CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
151    GGCAAACTGT ATGCGGAAgc CAAATTCGCC GACgGcAGCG TAACTTACGG
201    CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTTGT
251    TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCCGGGG
301    CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
351    GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAAACC GAAGTCGTCA
401    AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT cTTCGCACCG
451    TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC
501    CTATACGCTG AAACTCAAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC
551    CGTAA
```

Number 93 ORF

```
  1  ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
 51  .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
101  CTCCGTTTGC GGTTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
151  GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
201  GATGGTGTTT TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATCGTCC
251  CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
301  ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351  CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
401  ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
451  AGGCAGGGCG GCAATATT..
```

Number 94 ORF

```
  1  ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 51    TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
101    TTTGCACGTC CTGCCCGCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
151    CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
201    TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
251    ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCTGTGC
301    AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
351    TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
401    GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
451    GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
501    CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..
```

Number 95 ORF

```
  1  ..GCCGGCGCGA GTGCGAACAA CATTTCCGCG CGTTTTGCGG AAACACCCGT
```

51 CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG
101 TTACCGAATA TGAAAACTTC CTGCTGCTTA TCGGCTCGGT ATTTGCGCCG
151 ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAAC GGCGTGA

Number 96 ORF

1 ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
51 GTTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACTTTTC GATAAAAGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TAGCCGCCGC CATGCTCGCG
151 CCTGCAGCGG A.ACGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
201 GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
251 CGATGATGCA GGAAAACGGC AGCCTGATTG TATGGCACGG GCAGGACAAG
301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGT.ACGGA
351 TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTAA GACGGCATCT ACCTGCCGAC CGAAGC.CAG
451 CTCGACGGGC GGCAATTATA GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GCCTGCAAG..

Number 97 ORF

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGtCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
301 AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
351 TGAAAATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT
401 GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
451 GTAG

Number 98 ORF

1 ..GTGTCGCTGG CTTCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
51 CAACCAAATG CGGAAAACCC GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
101 ATTTATCTGG GGTTTCAGCA GGGGTATTTC GATTTGAGTG CCGACGAGAA
151 CCCCGTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGTATCCCCT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG
251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTTGGCAGG TTCGCTGCTG
401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTCGG CAAGCCCCAT CGTATTTGTC GGCAAAATCT
651 CTTATTCCCT ATACCTGTAC CATTGGATTT TTATTGCTTT CGCTCCGCTC
701 ATTAGAGGCG GGAAACAGCT CGGACTGCCT GCCG..

Number 99 ORF

1 ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51 GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
101 TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCGC
151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTTGGGCG TATGTGTGGT
251 TTCCGTTTTT CGTC..

Number 100 ORF

```
  1   ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
 51     TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
101     GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151     CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACTCT TACGTAAACA
201     cTACGTCCGC ACTTATTACy TGCTCCAACT CTTTGCCGCC GCAGgcTAgT
251     TTGTGGACAG GCGCGGCGwA ATTACAAAAC CTGCCCGCyT CCGCGCCCCT
301     GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
351     TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACTCGA CTACCCCAAA
401     CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451     TTTCTTGrTG AACGTGAACC CGrTATTTTT CATTACCGTT CCTGCGATTC
501     TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATTT
551     CGGGCGAATG CGTTTACAGA CGATCCGGAr TAr
```

Number 101 ORF

```
  1   ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
 51   TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101   CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
151   GGCGGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201   CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251   ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301   ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
351   CTGCTTGGAA AAG..
```

Number 102 ORF

```
  1   ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
 51   TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101   AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151   TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201   CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251   TGAACCTCGG CCTGCCCtTAT ATtTcCGGCC CGCAATGGCT GTCGGAAAAC
301   GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
351   GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
401   CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATttCGGCG TTTCCGCCCG
451   CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
501   TCATCGAAGC CGACGAATAC GACACCGCCT TTtTCGACAA ACGTTCTAAA
551   TtCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAATTCGA
601   CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACc CAGTTCCACT
651   ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
701   CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGA
751   AAAATTCGGC ACGGAACACG GCTGGCA..
```

Number 103 ORF

```
  1   ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
 51     CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
101     CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
151     ATTAGTGCGG ACTTCGGCGA TTATTTCATG CCGTTCGCCA GCTATTCGCG
201     CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTCC CAAATCGGCG
251     ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAAA CACTTGGCAA
301     TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAAAACAAG ATGATACATT
351     AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
401     ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
451     AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCAwAGACAA
501     AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
551     TTTTCACCAA CCTTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
601     TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
651     CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
701     GACGTTTGGA AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
```

751 GGCGCGATGC GCTATTTCGG CAAGAGCATC CGCGCGACGG CTGAAGAACG
801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
851 GCAAGCGTTC CATCAAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
901 TTwGATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
951 CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG
1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
1151 CCTTTTTgAT GACGATGAGC TACAAGTTTT AA

Number 104 ORF

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

Number 105 ORF

1 ATGAAACTTC TGACCACCGC AATCCTGTCT TCCGCAATCG CGCTCAGCAG
51 TATGGCTGCC GCCGCTGGCA CGGACAACCC CACTGTTGCA AAAAAAACCG
101 TCAGCTACGT CTGCCAGCAA GGTAAAAAAG TCAAAGTAAC CTACGGCTTC
151 AACAAACAGG GTCTGACCAC ATACGCTTCC GCCGTCATCA ACGGCAAACG
201 CGTGCAAATG CCTGTCAATT TGGACAAATC CGACAATGTG GAAACATTCT
251 ACGGCAAAGA AGGCGGTTAT GTTTTGGGTA CCGGCGTGAT GGATGGCAAA
301 TCCTACCGCA AACAGCCCAT TATGATTACC GCACCTGACA ACCAAATCGT
351 CTTCAAAGAC TGTTCCCCAC GTTAA

Number 106 ORF

1 ..ACACTGTTGT TTGCAACGGT TCAGGCAAGT GCTAACCAAT GAAGAGCAAG
51 AAGAAGATTT ATATTTAGAC CCCGTACAAC GCACTGTTGC CGTGTTGATA
101 GTCAATTCCG ATAAAGAAGG CACGGGAGAA AAAGAAAAAG TAGAAGAAAA
151 TTCAGATTGG GCAGTATATT TCAACGAGAA AGGAGTACTA ACAGCCAGAG
201 AAATCACCyT CAAAGCCGGC GACAACCTGA AAATCAAACA AAACGGCACA
251 AACTTCACCT ACTCGCTGAA AAAAGACCTC AcAGATCTGA CCAGTGTTGG
301 AACTGAAAAA TTATCGTTTA GCGCAAACGG CAATAAAGTC AACATcACAA
351 GCGACACCAA AGGCTTGAAT TTTGCGAAAG AAACGGCTGG sACGAACGgC
401 GACACCACGG TTCATCTGAA CGGTATTGGT TCGACTTTGA CCGATACGCT
451 GCTGAATACC GGAGCGACCA CAAACGTAAC CAACGACAAC GTTACCGATG
501 ACGAGAAAAA ACGTGCGGCA AGCGTTAAAG ACGTATTAAA CGCTGGCTGG
551 AACATTAAAG GCGTTAAACC CGGTACAACA GCTTCCGATA ACGTTGATTT
601 CGTCCGCACT TACGACACAG TCGAGTTCTT GAGCGCAGAT ACGAAAACAA
651 CGACTGTTAA TGTGGAAAGC AAAGACAACG GCAAGAAAAC CGAAGTTAAA
701 ATCGGTGCGA AGACTTCTGT TATTAAAGAA AAAGAC...

Number 107 ORF

1 ..GGCACCGAAT TCAAAACCAC CCTTTCCGGA GCCGACATAC AGGCAGGGGT
51 GGGTGAAAAA GCCCGAGCCG ATGCGAAAAT TATCCTAAAA GGCATCGTTA
101 ACCGCATCCA AACCGAAGAA AAGCTGGAAT CCAACTCGAC CGTATGGCAA
151 AAGCAGGCCG GAAGCGGCAG CACGGTTGAA ACGCTGAAGC TACCGAGCTT
201 TGAAGGGCCG GCACTGCCTA AGCTGACCGC TCCCGGCGGC TATATCGCCG
251 ACATCCCCAA AGGCAACCTC AAAACCGAAA TCGAAAAGCT GGCCAAACAG
301 CCCGAATATG CCTATCTGAA ACAGCTTCAG ACGGTCAAGG ACGTGAACTG

```
351    GAACCAAGTA CAGCTCGCTT ACGACAAATG GGACTATAAA CAGGAAGGCC
401    TAACCGGAGC CGGAGCCGCA ATTANCGCAC TGGCCGTTAC CGTGGTCACC
451    TCAGGCGCAG GAACCGGAGC CGTATTGGGA TTAANACGNG TGGCCGCCGC
501    CGCAACCGAT GCAGCATTT...
```

Number 108 ORF

```
  1  ..CGGATCGTTG TAGGTTTGCG GATTTCTTGC GCCGTAGTCA CCGTAGTCCC
 51    AAGTATAACC CAAGGCTTTG TCTTCGCCTT TCATTCCGAT AAGGGATATG
101    ACGCTTTGGT CGGTATAGCC GTCTTGGGAA CCTTTGTCCA CCCAACGCAT
151    ATCTGCCTGC GGATTCTCAT TGCCGCTTCT TGGCTGCTGA TTTTTCTGCC
201    TTCGCGTTTT TCAACTTCGC GCTTGAGGGC TTCGGCATAT TTGTCGGCCA
251    ACGCCATTTC TTTCGGATGC AGCTGCCTAT TGTTCCAATC TACATTCGCA
301    CCCACCACAG CACCACCACT ACCACCAGTT GCATAG
```

Number 109 ORF

```
   1  ..AAGTTTGACT TTACCTGGTT TATTCCGGCG GTAATCAAAT ACCGCCGGTT
  51    GTTTTTTGAA GTATTGGTGG TGTCGGTGGT GTTGCAGCTG TTTGCGCTGA
 101    TTACGCCTCT GTTTTTCCAA GTGGTGATGG ACAAGGTGCT GGTACATCGG
 151    GGATTCTCTA CTTTGGATGT GGTGTCGGTG GCTTTGTTGG TGGTGTCGCT
 201    GTTTGAGATT GTGTTGGGCG GTTTGCGGAC GTATCTGTTT GCACATACGA
 251    CTTCACGTAT TGATGTGGAA TTGGGCGCGC GTTTGTTCCG GCATCTGCTT
 301    TCCCTGCCTT TATCCTATTT CGAGCACAGA CGAGTGGGTG ATACGGTGGC
 351    TCGGGTGCGG GAATTGGAGC AGATTCGCAA TTTCTTGACC GGTCAGGCGC
 401    TGACTTCGGT GTTGGATTTG GCGTTTTCGT TTATCTTTCT GGCGGTGATG
 451    TGGTATTACA GCTCCACTCT GACTTGGGTG GTATTGGCTT CGTTG.....
                                  //
1451    .......... .......... .......... .......... ..........
1501    .......... .......... .......... .......... ..ATTTGCGC
1551    CAACCGGACG GTGCTGATTA TCGCCCACCG TCTGTCCACT GTTAAAACGG
1601    CACACCGGAT CATTGCCATG GATAAAGGCA GGATTGTGGA AGCGGGAACA
1651    CAGCAGGAAT TGCTGGCGAA CG..AACGGA TATTACCGCT ATCTGTATGA
1701    TTTACAGAAC GGGTAG
```

Number 110 ORF

```
  1  ATGAAATACT TGATCCGCAC CGCCTTACTC GCAGTCGCAG CCGCCGGCAT
 51  CTACGCCTGC CAACCGCAAT CCGAAGCCGC AGTGCAAGTC AAGGCTGAAA
101  ACAGCCTGAC CGCTATGCGC TTAGCCGTCG CCGACAAACA GGCAGAGATT
151  GACGGGTTGA ACGCCCAAAk sGACGCCGAA ATCAGA...
```

Number 111 ORF

```
  1  ATGGTTATCG GAATATTACT CGCATCAAGC AAGCATGCTC TTGTCATTAC
 51  TCTATTGTTA AATCCCGTCT TCCATGCATC CAGTTGCGTA TCGCGTTsGG
101  CAATACGGAA TAAAAtCTGC TGTTCTGCTT TGGCTAAATT TGCCAAATTG
151  TTTATTGTTT CTTTAGGaGC AGCTTGCTTA GCCGCCTTCG CTTTCGACAA
201  CGCCCCCACA GGCGCTTCCC AAGCgTTGCC TACCGTTACC GCACCCGTGG
251  CGATTCCCGC GCCCGCTTCG GCAGCCTGA
```

Number 112 ORF

```
  1  ATGTTCAGTA TTTTAAATGT GTTTCTTCAT TGTATTCTGG CTTGTGTAGT
 51  CTCTGGTGAG ACGCCTACTA TATTTGGTAT CCTTGCTCTT TTTTACTTAT
101  TGTATCTTTC TTATCTTGCT GTTTTTAAGA TTTTCTTTTC TTTTTTCTTA
151  GACAGAGTTT CACTCCGGTC TCCCAGGCTG GAGTGCAAAT GGCATGACCC
201  TTTGGCTCAC TGGCTCACGG CCACTTCTGC TATTCTGCCG CCTCAGCCTC
251  CAGGG...
```

Number 113 ORF

```
  1 ..GTGCGGACGT GGTTGGTTTT TTGGTTGCAG CGTTTGAAAT ACCCGTTGTT
 51   GCTTTGGATT GCGGATATGT TGCTGTACCG GTTGTTGGGC GGCGCGGAAA
101   TCGAATGCGG CCGTTGCCCT GTGCCGCCGA TGACGGATTG GCAGCATTTT
151   TTGCCGGCGA TGGGAACGGT GTCGGCTTGG GTGGCGGTGA TTTGGGCATA
201   CCTGATGATT GAAAGTGAAA AAAACGGAAG ATATTGA
```

Number 114 ORF

```
  1 ATGTTTCAAA ATTTTGATTT GGGCGTGTTC CTGCTTGCCG TCCTCCCCGT
 51 GCTGCCCTCC ATTACCGTCT CGCACGTGGC GCGCGGCTAT ACGGCGCGCT
101 ACTGGGGAGA CAACACTGCC GAACAATACG GCAGGCTGAC ACTGAACCCC
151 CTGCCCCATA TCGATTTGGT CGGCACAATC ATCgTACCGC TGCTTACTTT
201 GATGTTCACG CCCTTCCTGT TCGGCTGGGC GCGTCCGATT CCTATCGATT
251 CGCGCAACTT CCGCAACCCG cGCCTTGCCT GGCGTTGCGT TGCCGCGTCC
301 GGCCCGCTGT CGAATCTAGC GATGGCTGTw CTGTGGGGCG TGGTTTTGGT
351 GCTGACTCCG TATGTCGGCG GGGCGTATCA GATGCCGTTG GCTCAAATGG
401 CAAACTACGG TATTCTGATC AATGCGATTC TGTTCGCGCT CAACATCATC
451 CCCATCCTGC CTTGGGACGG CGGCATTTTC ATCGACACCT TCCTGTCGGC
501 GAAATATTCG CAAGCGTTCC GCAAAATCGA ACCTTATGGG ACGTGGATTA
551 TCCTACTGCT GATGCTGACC sGGGTTTTGG GTGCGTTTAT wGCACCGATT
601 sTGCGGmTGc GTGATTGCrT TTGTGCAGAT GTwCGTCTGA CTGGCTTTCA
651 GACGGCATAA
```

Number 115 ORF

```
  1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
 51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..
```

Number 116 ORF

```
  1 ..GCAGTAGCCG AAACTGCCAA CAGCCAGGGC AAAGGTAAAC AGGCAGGCAG
 51   TTCGGTTTCT GTTTCACTGA AAACTTCAGG CGACCTTTGC GGCAAACTCA
101   AAACCACCCT TAAAACTTTG GTCTGCTCTT TGGTTTCCCT GAGTATGGTA
151   TTGCCTGCCC ATGCCCAAAT TACCACCGAC AAATCAGCAC CTAAAAACCA
201   GCAGGTCGTT ATCCTTAAAA CCAACACTGG TGCCCCCTTG GTGAATATCC
251   AAACTCCGAA TGGACGCGGA TTGAGCCACA ACCGCTA.TA CGCATTTGAT
301   GTTGACAACA AAGGGGCAGT GTTAAACAAC GACCGTAACA ATAATCCGTT
351   TGTGGTCAAA GGCAGTGCGC AATTGATTTT GAACGAGGTA CGCGGTACGG
401   CTAGCAAACT CAACGGCATC GTTACCGTAG GCGGTCAAAA GGCCGACGTG
451   ATTATTGCCA ACCCCAACGG CATTACCGTT AATGGCGGCG GCTTTAAAAA
501   TGTCGGTCGG GGCATCTTAA CTACCGGTGC GCCCCAAATC GGCAAAGACG
551   GTGCACTGAC AGGATTTGAT GTGCGTCAAG GCACATTGgA CCGTAGrAGC
601   AGCAGGTTGG AATGATAAAG GCGGAGCmrm yTACACCGGG GTACTTGCTC
651   GTGCAGTTGC TTTGCAGGGG AAATTwmmGG GTAAA.AACT GGCGGTTTCT
701   ACCGGTCCTC AGAAAGTAGA TTACGCCAGC GGCGAAATCA GTGCAGGTAC
751   GGCAGCGGGT ACGAAACCGA CTATTGCCCT TGATACTGCC GCACTGGGCG
801   GTATGTACGC CGACAGCATC ACACTGATTG CCAATGAAAA AGGCGTAGGC
851   GTCTAA
```

Number 117 ORF

```
   1  ..CGCTTCATTC ATGATGAAGC AGTCGGCAGC AACATCGGCG GCGGCAAAAT
  51    GATTGTTGCA GCCGGGCAGG ATATCAATGT ACGCGGCAnA AGCCTTATTT
 101    CTGATAAGGG CATTGTTTTA AAAGCAGGAC ACGACATCGA TATTTCTACT
 151    GCCCATAATC GCTATACCGG CAATGAATAC CACGAGAGCA wAAAwTCAGG
 201    CGTCATGGGT ACTGGCGGAT TGGGCTTTAC TATCGGTAAC CGGAAAACTA
 251    CCGATGACAC TGATCGTACC AATATTGTsC ATACAGGCAG CATTATAGGC
 301    AGCCTGAaTG GAGACACCGT TACAGTTGCA GGAAACCGCT ACCGACAAAC
 351    CGGCAGTACC GTCTCCAGCC CCGAGGGGCG CAATACCGTC ACAGCCAAAw
 401    GCATAGATGT AGAGTTCGCA AACAACCGGT ATGCCACTGA CTACGcCCAT
 451    ACCCAgGGAA CAAAAAGGCC TTACCGTCGC CCTCAATGTC CCGGTTGTCC
 501    AAGCTGCACA AAACTTCATA CAAGCAGCCC AAAATGTGGG CAAAAGTAAA
 551    AATAAACGCG TTAATGCCAT GGCTGCAGCC AATGCTGCAT GGCAGAGTTA
 601    TCAAGCAACC CAACAAATGC AACAATTTGC TCCAAGCAGC AGTGCGGGAC
 651    AAGGTCAAAA CTACAATCAA AGCCCCAGTA TCAGTGTGTC CATTAC.TAC
 701    GGCGAACAGA AAAGTCGTAA CGAGCAAAAA AGACATTACA CCGAAgCGGC
 751    AgCAAGTCAA ATTATCGGCA AAGGGCAAAC CACACTTGCG GCAACAGGAA
 801    GTGGGGAGCA GTCCAATATC AATATTACAG GTTCCGATGT CATCGGCCAT
 851    GCAGGTACTC C.CTCATTGC CGACAACCAT ATCAGACTCC AATCTGCCAA
 901    ACAGGACGGC AGCGAGCAAA GCAAAAACAA AAGCAGTGGT TGGAATGCAG
 951    GCGTACGTnn CAAAATAGGC AACGGCATCA GGTTTGGAAT TACCGCCGGA
1001    GGAAATATCG GTAAAGGTAA AGAGCAAGGG GGAAGTACTA CCCACCGCCA
1051    CACCCATGTC GGCAGCACAA CCGGCAAAAC TACCATCCGA AGCGGCGGGG
1101    GATACCACCC TCAAAGGTGT GCAGCTCATC GGCAAAGGCA TACAGGCAGA
1151    TACGCGCAAC CTGCATATAG AAAGTGTTCA AGATACTGAA ACCTATCAGA
1201    GCAAACAGCA AAACGGCAAT GTCCAAGTTt ACTGTCGGTT ACGGATTCAG
1251    TGCAAGCGGC AGTTACCGCC AAAGCAAAGT CAAAGCAGAC CATGCCTCCG
1301    TAACCGGGCA AAgCGGTATT TATGCCGGAG AAGACGGCTA TCAAATyAAA
1351    GTyAGAGACA ACACAGACCT yAAGGGCGGT ATCATCACGT CTAGCCAAAG
1401    CGCAGAAGAT AAGGGCAAAA ACCTTTTTCA GACGGCCACC CTTACTGCCA
1451    GCGACATTCA AAACCACAGC CGCTACGAAG GCAGAAGCTT CGGCATAGGC
1501    GGCAGTTTCG ACCTGAACGG CGGCTGGGAC GGCACGGTTA CCGACAAACA
1551    AGGCAGGCCT ACCGACAGGA TAAGCCCGGC AGCCGGCTAC GGCAGCGACG
1601    GAGACAGCAA AAACAGCACC ACCCGCAGCG GCGTCAACAC CCACAACATA
1651    CACATCACCG ACGAAGCGGG ACAACTTGCC CGAACAGGCA GGACTGCAAA
1701    AGAAACCGAA GCGCGTATCT ACACCGGCAT CGACACCGAA ACTGCGGATC
1751    AACACTCAGG CCATCTGAAA AACAGCTTCG AC...
```

Number 118 ORF

```
   1  ..ACGACCGGCA GCCTCGGCGG CATACTGGCC GGCGGCGGCA CTTCCCTTGC
  51    CGCACCGTAT TTGGACAAAG CGGCGGAAAA CCTCGGTCCG GCGGGCAAAG
 101    CGGCGGTCAA CGCACTGGGC GGTGCGGCCA TCGGCTATGC AACTGGTGGT
 151    AGTGGTGGTG CTGTGGTGGG TGCGAATGTA GATTGGAACA ATAGGCAGCT
 201    GCATCCGAAA GAAATGGCGT TGGCCGACAA ATATGCCGAA GCCCTCAAGC
 251    GCGAAGTTGA AAAACGCGAA GGCAGAAAAA TCAGCAGCCA AGAAGCGGCA
 301    ATGAGAATCC GCAGGCAGAT ATGCGTTGGG TGGACAAAGG TTCCCAAGAC
 351    GGCTATACCG ACCAAAGCGT CATATCCCTT ATCGGAATGA
```

Number 119 ORF

```
   1  ..CAATGCCGTC TGAAAAGCTC ACAATTTTAC AGACGGCATT TGTTATGCAA
  51    GTACATATAC AGATTCCCTA TATACTGCCC AGrkGCGTGC GTgGCTGAAG
 101    ACACCCCCTA CGCTTGCTAT TTGrAACAGC TCCAAGTCAC CAAAGACGTC
 151    AACTGGAACC AGGTACwACT GGCGTACGAC AAATGGGACT ATAAACAGGA
 201    AGGCTTAACC GGAGCCGGAG CAGCGATTAT TGCGCTGGCT GTTACCGTGG
 251    TTACTGCGGG CGCGGGAgCC GGAGCCGCAC TGGGcTTAAA CGGCGCGGCc
 301    GCAGCGGCAA CCGATGCCGC ATTCGCCTCG CTGGCCAGCC AGGcTTCCGT
 351    ATCGCTCATC AaCAACAAAG GCAATATCGG TAaCACCCTG AAAGAGCTGG
 401    GCAGAAGCAG CACGGTGAAA AATCTGATGG TTGCCGTCGc tACCGCAgGC
 451    GTagCcgaCA AAATCGGTGC TTCGGCACTG AACAATGTCA GCGATAAGCA
 501    GTGGATCAAC AACCTGACCG TCAACCTGGC CAATGCGGGC AGTGCCGCAC
```

```
551   TGATTAATAC CGCTGTCAAC GGCGGCAGCc tgAAAGACAA TCTGGAAGCG
601   AATATCCTTG CGGCTTTGGT GAATACTGCG CATGGAGAAG CAGCCAGTAA
651   AATCAAACAG TTGGATCAGC ACTACATTAC CCACAAGATT GCCCaTGCCA
701   TAGCGGGCTG TGCGGcTCCG GCGGCGAATA AGGGCAAGTG TCAGGATGGT
751   GCGATAgGTG CGGCTGTGGG CGAGATAGTC GGGGAgGCTT TGACAAACGG
801   CAAAAATCCT GACACTTTGA CAGCTAAAgA ACGCCGaACAG ATTTTGGCAT
851   ACAGCAAACT GGTTGCCGGT ACGGTAAGCG GTGTGGTCGG CGGCGATGTA
901   AATGCGGCGG CGAATGCGGC TGAGGTAGCG GTGAAAAATA ATCAGCTTAG
951   CGACAAAtGA
```

Number 120 ORF

```
  1   ATGGCAATCA TTACATTGTA TTATTCTGTC AATGGTATTT TAAATGTATG
 51   TGCAAAAGCA AAAAATATTC AAGTAGTTGC CAATAATAAG AATATGGTTC
101   TTTTTGGGTT TTTGGsmrGC ATCATCGGCG GTTCAACCAA TGCCATGTCT
151   CCCATATTGT TAATATTTTT GCTTAGCGAA ACAGAAAATA AAAATcgTAT
201   CGTAAAATCA AGCAATCTAT GCTATCTTTT GGCGAAAATT GTTCAAATAT
251   ATATGCTAAG AGACCAGTAT TGGTTATTAA ATAAGAGTGA ATACGdTTTA
301   ATATTTTTAC TGTCCGTATT GTCTGTTATT GGATTGTATG TTGGAATTCG
351   GTTAAGGACT AAGATTAGCC CAaATTTTTT TAAAATGTTA ATTTTTATTG
401   tTTTATTGGT ATTGGCtCTG AAAATCGGGC AttCGGGTTT AAtCAAACTT
451   TAA
```

Number 121 ORF

```
  1   ATGTTACGTt TGACTGCtTT AGCCGTATGC ACCGCCCTCG CTTTGGGCGC
 51   GTGTTCGCCG CAAAATTCCG ACTCTGCCCC ACAAGCCAAA GaACAGGCGG
101   TTTCCGCCGC ACAAACCGAA GgCGCGTCCG TTACCGTCAA AACCGCGCGC
151   GGCGACGTTC AAATACCGCA AAACCCCGAA CGCATCGCCG TTTACGATTT
201   GGGTATGCTC GACACCTTGA GCAAACTGGG CGTGAAAACC GGTTTGTCCG
251   TCGATAAAAA CCGCCTGCCG TATTTAGAGG AATATTTCAA AACGACAAAA
301   CCTGCCGGCA CTTTGTTCGA GCCGGATTAC GAAACGCTCA ACGCTTACAA
351   ACCGCAGCTC ATCATCATCG GCAGCCGCGC CgCCAAGGCG TTTGACAAAT
401   TGAAcGAAAT CGCGCCGACC ATCGrmwTGA CCGCCGATAC CGCCAACCTC
451   AAAGAAAGTG CCAArGAGGC ATCGACGCTG GCGCAAATCT TC..
```

Number 122 ORF

```
  1   ATGAGACATA TGAAAATACA AAATTATTTA CTAGTATTTA TAGTTTTACA
 51   TATAGCCTTG ATAGTAATTA ATATAGTGTT TGGTTATTTT GTTTTTCTAT
101   TTGATTTTTT TGCGTTTTTG TTTTTTGCAA ACGTCTTTCT TGCTGTAAAT
151   TTATTATTTT TAGAAAAAAA CATAAAAAAC AAATTATTGT TTTTATTGCC
201   GATTTCTATT ATTATATGGA TGGTAATTCA TATTAGTATG ATAAATATAA
251   AATTTTATAA ATTTGAGCAT CAAATAAAGG AACAAAATAT ATCCTCGATT
301   ACTGGGGTGA TAAAACCACA TGATAGTTAT AATTATGTTT ATGACTCAAA
351   TGGATATGCT AAATTAAAAG ATAATCATAG ATATGGTAGG GTAATTAGAG
401   AAACACCTTA TATTGATGTA GTTGCATCTG ATGTTAAAAA TAAATCCATA
451   AGATTAAGCT TGGTTTGTGG TATTCATTCA TATGCTCCAT GTGCCAATTT
501   TATAAAATTT GTCAGG..
```

Number 123 ORF

```
  1   ..ACCCCCAACA GCGTGACCGT CTTGCCGTCT TTCGGCGGAT TCGGCCGTAC
 51     CGGCGCGACC ATCAATGCAG CAGGCGGGGT CGGCATGACT GCCTTTTCGA
101     CAACCTTAAT TTCCGTAGCC GAGGGCGCGG TTGTAGAGCT GCAGGCCGTG
151     AGAGCCAAAG CCGTCAATGC AACCGCCGCT TGCATTTTTA CGGTCTTGAG
201     TAAGGACATT TTCGATTTCC TTTTTATTTT CCGTTTTCAG ACGGCTGACT
251     TCCGCCTGTA TTTTCGCCAA AGCCATGCCG ACAGCGTCCG CCTTGACTTC
301     ATATTTAAAA GCTTCCGCGC GTGCCAGTTC CAGTTCGCGC GCATAGTTTT
351     GAGCCGACAA CAGCAGGGCT TGCGCCTTGT CGCGCTCCAT CTTGTCGATG
401     ACCGCCTGCA GCTTCGCAAA TGCCGACTTG TAGCCTTGAT GGTGCGACAC
```

```
451   AGCCAAGCCC GTGCCGACAA GCGCGATAAT GGCAATCGGT TGCCAGTAAT
501   TCGCCAGCAG TTTCACGAGA TTCATTCTCG ACCTCCTGAC GCTTCACGCT
551   GA
```

APPENDIX C

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1>:

gnm_1

GAAAATTCAGCAGCAGCAGGAAGATTGCCAGCATTTGCGCGGCGGTTTTAAACTTACCGA
CGGTGGCGACGGCAACGCTGTTCCTTTTGCCCATTTGCGCCATCCATTCGCGCAATGCGG
AAATGGTAATTTCCCTGCCGATGATGATCATGGCAAACAAAACATAGGTCCGGTCGAGTT
TGACCAGTAAAAGCAAAGAGACGGCGACCATCAGCTTGTCGGCAACGGGATCGAGGAAGG
CGCCGAAATCCGAGGTCTGTTTCCACAACCTTGCCAAAAATCCGTCAAACCAGTCGGTCA
AGGCGGCAACGGCAAAAATGACGGCGGCGGTGAGATTAATCGTTTCCTCCGCGAACCACG
GAAAAGGCAGGTAAAAAAGGGCTGTCAGGACAGGAATGAGCAAGACCCTCAACCATGTGA
GGAAGATGGGGAGATTCCAAGGCATCGGTTTTCTCTGTGCAGACTGTAAAGTTGTGATTA
TAACGGTTATCCTCATAACCCAAAACGTAAAATTGCTGCATGGGCATTCCCCCGCCCCGC
CAATCTGTTTTCACATTCTTTTCAAACGCAGGAAAATGGCGGGCAATAAAAGCAAAATAC
CCAGTTTCAGGCTGAAAACGGCAGGTTGTGCCAACACTTCGACAAGGCGGTCTTCCGTGC
GGGCAAAATCTTTATTGCTTATAGACACTGCCACTGTTGCGGTATTCCAACAGAACGCCG
TTTAAAAAACCTTTGCCGACGGTTTCGCTTAAAACGGCTCTAACCTGCTCCGCCCTGATG
GTTCTGCCGATATTGCCGCCTGTGCACAAACTGTCGAACCCATAGCAGGAAAGCCGGTAA
TGCTGCCCGTCTGCATCCAGTTTGATTGCCCGTCCGCTGCGGTTGAGGGCGGTAACGGTC
AATTCCGCATATTCGAATGTTTTTTCTTGTTCGTGAAATGCCGTCAGGTAAGGTGCAATA
AAAACGGCGGACAACAGCAGACAGCTTATGGCGGCAAACCATACCCAGCGATAATATAGT
GGATTAAATTTAAACCAGTACAGCGTTGCCTCGCCTTAGCTCAAAGAGAACGATTCTCTA
AGGTGCTGAAGCACCAAGTGAATCGGTTCCGTACTATTTGTACTGTCTGCGGCTTCGTCG
CCTTGTCCTGATTTAAATTTAATCCACTATATTTCACGCTTACCCCTTGTTTCTCAAATG
CCGTCTGAAATAAGCGGCTTAATATATTGTTTACAGTATTGGGAAGCATAACAGACAAAA
TGCCGTCTGAAATATTTTCAGACGGCATTTCTTATCCGAAACGGATTATTTTTGCGTTTC
AACCGCTTCCAATGCACGCAGGGCATAAGTGTAAGCGGCACCCGCATTCAGGGCAATGGC
GGTTGCCAATGCACCTGCGATTTCGCTGTCGGTCGCACCGGCTTTGGTGGCGGCGGCGGC
GTGAACACTGATGCAGCTCTCACAACGTGTAGTAATGGCAACGGCGATGGCAATCAGTTC
GCGTGTTTTAGCATCAAGTGCCTCTGCAGCTGCCGCTTGTTCCAATGCGCCGTAGGCCTG
CAGCATTTTAGGATGCGCCTTACCCAGCTCGCCGAACGATTTTTTAACCAATGCGGTATG
TTCTTTCCAATCTTTAAACATTTTCTTTTCCTTTCTCTTGCGTTTAACCCTGATACGCGC
TTGCGTATCTGTTTTCGATGTGCGTATTATTGCAATTATTCAGTTGTGTTTCTCGTTTAA
TCATCTCATTTTATGGTTCAAAAAGATTTATGGACATTCTGGACAAACTGGTCGATTTCG
CCCAATTGACGGGCAGTGTGGATGTGCAGTGCCTTTTGGGCGGACAATGGTCGGTACGGC
ATGAAACCTTGCAACGCGAAGGATTGGTACACATTGTTACATCGGGCAGCGGCTATCTCT
GCATCGACGGCGAAACTTCCCCGCGTCCGGTCAGTACAGGGGATATTGTATTTTTCCCGC
GCGGCTTGGGTCATGTGTTGAGCCACGACGGAAAATGCGGAGAAAGTTTACAACCGGATA
TGCGGCAGCACGGTGCGTTTACGGTCAAGCAGTGCGGCAACGGACAGGATATGAGCCTGT
TTTGCGCCCGTTTCCGCTACGACACCCACGCCGATTTGATGAACGGGCTGCCTGAAACCG
TTTTTCTGAACATTGCCCATCCGAGTTTACAGTATGTGGTTTCAATGCTGCAACTGGAAA
GCAAAAAACCTTTGACGGGGACGGTTTCCATGGTCAACGCATTGTCGTCCGTCCTGCTGG
TGCTTATCCTGCGCGCCTATCTCGAACAGGATAAGGATGTCGAACTCTCGGGCGTATTGA
AAGGTTGGCAGGACAAACGTTTGGGACATTTAATCCAAAAGGTGATAGACAAACCGGAAG
ACGAATGGAATGTCGACAAAATGGTGGCGGCTGCCAATATGTCGCGCGCGCAACTGATGC
GCCGTTTCAAAAGCCGGGTCGGACTCAGCCCGCACGCCTTTGTGAACCATATCCGCCTGC
AAAAAGGCGCGTTGCTGCTGAAAAAAAACCCGGATTCGGTTTTGTCGGTCGCACTGTCGG
TAGGCTTTCAGTCGGAAACGCACTTCGGCAAGGCGTTCAAACGGCAATATCACGTTTCGC
CGGGTCAATACCGGAAAGAAGCGGGCAAAAATAAATCGGGGCTTCAAACGCAAATGCCGT
CTGAAAAGGCTTTCATACAGCATTTGCGTACCGCGTCATTTCAAGGGCTGCATCTTCATC
ACTTCCATCAAAAAGTTGGTAAATGCGGGGTTGTTGGGTTTGACATCCATATTTTTCCAA
CGCTGCTGCCAGCCGCGCAAGGCATTCTGGATATACAGCTTGGACTGTTCCGTATTGATT
GCGCCCCGCTGGCTGTCTATCGCCGAACGCAGGTAGATTTCATACATACTGTCATCGACG
GCATTGCGTCCGACCAGGCGTTTTCTGAAGTTGTTCAGATATTGCGCCGCCTGAACCTTG

GTCATTTTACCGATACCCACCTGATAGCCCAAGCGCGTCGCTTCATCGCTGATTTTGGCA
ACATCCGTCCAATGCGAAGAGGCAAGGCGGAAACCTTTTGCAGGTGCTTCCGTTTTGACG
GTATTGATAGGATTCACGGGGATTTCCGTCAATGTGGGCACATAAATAGACTGGCAGCCG
GAAAGAACTGCCGCAATGGAAAGAGGGATAAGGTATTTTTTCATGCCCCCATTATAATCA
AGTTTGCCTTGAGAAAACAAATTGTTCGGCAAGAAAAATAAAATTTCGGCATCAGAAGCA
GGCAAAAACACATICCACAAGCCTTGCCGCAAGGTTTACAATCCGACCGTCCTTATCGCA
ACGACCGTTTATGGATACCGCAAAAAAAGACATTTTAGGATCGGGCTGGATGCTGGTGGC
GGCGGCCTGCTTTACCATTATGAACGTATTGATTAAAGAGGCATCGGCAAAATTTGCCCT
CGGCAGCGGCGAATTGGTCTTTTGGCGCATGCTGTTTTCAACCGTTGCGCTCGGGGCTGC
CGCCGTATTGCGTCGGGACAmCTTCCGCACGCCCCATTGGAAAAACCACTTAAACCGCAG
TATGGTCGGGACGGGGGCGATGCTGCTGCTGTTTTACGCGGTAACGCATCTGCCTTTGGC
CACTGGCGTTACCCTGAGTTACACCTCGTCGATTTTTTTGGCGGTATTTTCCTTCCTGAT
TTTGAAAGAACGGATTTCCGTTTACACGCAGGCGGTGCTGCTCCTTGGTTTTGCCGGCGT
GGTATTGCTGCTTAATCCCTCGTTCCGCAGCGGTCAGGAAACGGCGGCACTCGCCGGGCT
GGCGGGCGGCGCGATGTCCGGCTGGGCGTATTTGAAAGTGCGCGAACTGTCTTTGGCGGG
CGAACCCGGCTGGCGCGTCGTGTTTTACCTTTCCGTGACAGGTGTGGCGATGTCGTCGGT
TTGGGCGACGCTGACCGGCTGGCACACCCTGTCCTTTCCATCGGCAGTTTATCTGTCGTG
CATCGGCGTGTCCGCGCTGATTGCCCAACTGTCGATGACGCGCGCCTACAAAGTCGGCGA
CAAATTCACGGTTGCCTCGCTTTCCTATATGACCGTCGTTTTTTCCGCTCTGTCTGCCGC
ATTTTTTCTGGGCGAAGAGCTTTTCTGGCAGGAAATACTCGGTATGTGCATCATCATCCT
CAGCGGTATTTTGAGCAGCATCCGCCCCACTGCCTTCAAACAGCGGCTGCAATCCCTGTT
CCGCCAAAGATAAAAAATGCCGTCCGAACATCCTTCAGACGGCATATCGGGCTTTATTTC
CCCGCCTTCACATCCTGCCACTGGCGCACCATAAACTTCAATGCCGCCGGCTGGATAGGC
ACCATGATAAAGCTGTTTTTCAAATCCTCCTCGGTTGGGAAAATCGTATTGTCGTTTTTA
AATTCGTCTTCCATCAGCTCACGCGCAGGCTTGCTCGAAGGCGCGTAAGTAACGAAATTG
CCGTTTTTCGCCGACACTTCCGGGTCGAGGAAGTCGTTGATGTATTTGTGCGCGTTGGCG
ACGTTTTTCGCATCTTTCGGAATCACGAAAGAATCCACCCAAATCCCCACGCCCTCTTTG
GGCATCATCACGCGGATTTTTTCCTTGCCGCCCGCTTCTTCGGCACGGCGTTTGGCGATG
TTCAAATCGCCGCCGAAACCGATTGTTACGCAGGTATCGCCGCGCGCCAAATCATCGATA
AAGCCGGACGAAGTAAAGCGTTTGATATTGGGGCGGTTTTTCTTGAGTAGGGCGGTTGCC
TCCCTGATGTCTTCCGTATTGCTGCTGTTCGGGTTTTTACCCAAATAGTTCAACACCATA
GGATAGATTTCCGCCGCGCTGTCCAAATAGCTGATGCCGCATTGCTTGAGTTTGGACGTG
TATTCGGGGTCGAACACCAAATCCCACTGGTTGTCCGGCAGCTTGTCCGTACCCAAAGCC
TTTTTCACGCGTTCGGTATTGATGGCGAAGGTATTTGTCCCCCAATAAAACGGCACGGCG
TATTCGTGGCCGGGATCGACCCCGTCCATCAGCCTCATCATTTCGGGGTTGAGGTGTTTA
TAATTGGGAATCAGCGACTTATCGATTTTCTGATACGCACCTGCCTTAATCTGCCTGCCC
ACAAACGCATTGGACGGCGCGACAATGTCGTAACCGGACTTGCCTGTCAGCACCTTGCTT
TCCAGCGTTTCATCGCTGTCGTACACATCATAAGTAACCTTGATGCCGTTTTTCTTTTCA
AAATCGGCAACGGTTTCCGGATCGACATATTCCGACCAGTTGTAAATTTTCAATACGTTT
TGGTTTTCCGCCGGTGCCGGTTTTTCGGCAGGCGGTTTGTCCGAACCGCCGCACGCTGCA
AGCAGCAAAGCAGTCAGGACGGCCAGGGGCAGATGTTTGGTCATTATCATTCCTTGCATA
TCGGGTTGGAGAAAGCGGCCATTATAGCCGATATTGGCAACAGGGCTTCAGACGGCATTC
AAAATCCCGCCACACTCTTCCGAAAACCGCCGCTTCCATAGCTAGAAACAGGGATTTGCG
GTAAGATACCGCCGTTCGTTTTCCCTGCTTTTACCATGACAAGACATTTGAGAGACATTG
AAAAAATTATGAAAACCTCCGAACTGCGCCAAAAATTCCTAAAATTTTTTGAAACCAAAG
GCCACACCGTCGTCCGCTCTTCCAGCCTCGTGCCGCACGACGACCCGACCCTGCTGTTTA
CCAACGCGGGCATGAACCAGTTTAAAGACGTATTCTTAGGTTTCGACAAACGCCCGTACA
GCCGCGCCACCACCGCGCAAAAATGCGTACGCGCAGGCGGCAAACACAACGACTTGGAAA
ACGTCGGCTACACCGCCCGCCACCACACCTTCTTTGAAATGATGGGCAACTTCTCCTTCG
GCGACTACTTCAAACGCGACGCCATCCACTTCGCTTGGGAATTTCTGACTTCCCCCGAAT
GGCTCAACATCCCTAAAGACAAACTGTTGGCGACCGTTTACGCGGAAGACGACGAAGCCT
ACAACATCTGGTTGAACGAAATCGGTATGCCGTCCGAGCGCATCGTCCGCATCGGCGACA
ACAAAGGCGCGAAATACGCATCCGACAACTTCTGGCAAATGGGCGACACCGGCCCTTGCG
GCCCCTGCTCCGAAATTTTCTACGACCACGGCGAAGAAATCTGGGGCGGCATTCCCGGCA
GTCCCGAAGAAGACGGCGACCGCTGGATCGAAATTTGGAACTGCGTATTTATGCAGTTCA
ACCGCGACGAACAAGGCAATATGAACCCGCTTCCCAAACCTTCCGTCGATACCGGTATGG
GCTTGGAACGCATAGCCGCCGTCATGCAGCATGTTCACAGCAACTACGAAATCGACTTGT

TCCAAGACCTGCTCAAAGCCGTTGCCCGCGAAACCGGCGCGCCGTTCAGAATGGAAGAAC
CCAGCCTGAAAGTCATCGCCGACCACATCCGCTCCTGCTCGTTCCTGATTGCAGACGGCG
TCTTGCCTTCCAACGAAGGCCGCGGCTACGTATTGCGCCGCATTATCCGCCGCGCCGTGC
GCCACGGTTACAAACTGGGTCAAAGCAAACCGTTCTTCCACAAACTCGTTGCCGATTTGG
TCAAAGAGATGGGCGGTGCCTACCCTGAATTGAAAGAAAAACAAGCCCAAATCGAAGAAG
CATTGAAAAACGAAGAAAGCCGTTTTGCCCAAACGCTGGAAACCGGTATGGCTTTGTTGG
AAAACGCGCTGGTyAAAGGCGGCAAAACACTCGGCGGCGAAATCATCTTCAAACTCTACG
ATACCTACGGTTTCCCATACGACTTGACTGCCGACATCTGCCGCGAACGCAATATCGAAC
CGGACGAAGCAGGCTTCGAGCGCGAAATGGAAGCCCAACGCGCACGCGCACGCGCCGCCC
AAAGCTTCAAAGCCAACGCCCAACTGCCTTATGACGGTCAAGACACCGAGTTTAAAGGTT
ATAGCGAACGCCAAACCGAATCCAAAGTCCTCGCCCTCTACAAAGACGGCGAGCAAGTCA
ACGAATTGAACGAAGGCGACAGCGGCGCAGTCGTCATCGACTTTACCCCGTTCTATGCAG
AATCCGGCGGCCAAGTCGGCGATGTCGGCTATATCTTCTCAGGCGAAAACCGCTTTGAAG
TACGCGATACCCAAAAAATCAAAGCGGCCGTATTCGGTCAATTCGGCGTACAAACTTCAG
GCCGTCTGAAAGTCGGCGACAGCGTTACCGCCAAAGTGGACGACGAAATCCGCAATGCCA
ATATGCGCAACCACAGCGCAACCCACTTGATGCACAAAGCCCTGCGCGATGTATTGGGCA
GACACGTCGAACAAAAAGGCTCTTTGGTTACCGCCGAATCCACCCGTTTCGACATTTCCC
ATCCCCAAGCGGTAACTGCCGAAGAAATTGCCGAAGTAGAACGCCGCGTCAACGAAGCCA
TTTTGGCGAACGTTGCCGTCAATGCAGCCATTATGAGCATGGAAGACGCGCAAAAAACCG
GCGCGATGATGCTCTTCGGCGAAAAATACGGCGAAGAAGTGCGCGTACTGCAAATGGGCG
GTTTCTCTACCGAATTGTGCGGCGGCACACACGTTTCACGCACCGGCGACATCGGCCTCT
TCAAAATCATCAGCGAAGGCGGTATTGCCGCAGGCGTGCGCCGTATCGAAGCCATCACCG
GCCTGAACGCACTCAAATGGGCGCAAGAGCAAGAGCGTTTGGTGAAAGACATTATTGCCG
AAACCAAAGCCCAAACCGAAAAAGACGTACTGGCAAAAATCCAAGCAGGCGCGGCACACG
CCAAAGCATTGGAAAAAGAATTGGCACGCGCCAAAGCCGAACTCGCCGTCCACGCAGGCG
CCAAACTCTTGGACGATGCAAAAGACTTGGGCGCAGCCAAACTCGTTGCCGCCCAAATCG
AAGCCGACGCAGCCGCCCTGCGCGAAATCGTTACCGATTTAACCGGTAAATCCGACAACG
CCGTGATTCTTTTAGCGGCAGTAAACGACGGCAAAGTCTCCCTGTGCGCCGGCGTATCCA
AACCGTTGACCGGCAAAGTGAAAGCAGGCGATCTGGTTAAATTTGCAGCCGAACAAGTCG
GCGGCAAAGGCGGCGGCAGACCAGATTTGGCGCAAGCCGGCGGCACGGATGCCGACAAAT
TGCCCGCCGTGTTGGATAGCGTGAAAGACTGGGTCGGCGCGAAGCTGGTTTGATGTGGGA
AAGGCAGCCTGAAAGGTTTCAGGCTGCCTTTTGTGCAAAGAGGCCGTCTGAAAGGTCTCG
TTTGCCGTAGGTTGGGTCGCGACCCAACAAATTTTGTGAAGTATAAAAATGTTGGTCATG
ACCCAACCTACCTGCCTTTTTGTACAAAGAGGCTATCTGAAAGGCCTTGTTTGCCGTATG
GTGGGTCGCGACCCAGCAGATTTTTATTAGGGTATGACCCAAGCTACTTGCTACGATAAA
AAAGGATTTTTAAATGAGCATTAGCCTTATTGGACTACACATTACCATAGCAATCATTTT
GTTTTTTACTACAAATTTTATGGGAAAAAAATCATCTATATTTGGCTATTACCAACTGTC
TTTTAGCGAAGAAAATCACTCTCCGGCATTTAATATTTTTTACAGAGCATTTACCCCTAT
ATTATTTATCGTTATTTTTTCTTGGGTTGTTACTAGTCTTGAAATTCCCATTTCTCTTGA
AAAGATAAACTATGTAGTAATTTATTATTTTATAATTAGATTGTTATCTGTATTTGTTTT
TGAGAAAACACACATAGTTAACTGGTTTAATCAACTAACAATACCCATACTATCCATAAC
ATTATCATTTATAGTATATAACAAAATGATTTTGCCCAAAAGTTTTCTACTTCCATCCTC
ACAAGAAGTAGCTACTACTTTTTGAATAGCGCTTGGTGGTTACATATATAATATATTAAA
TAATGAATCAGGGCATTTAAAATCTTATAAAGAAAGAAGAGTAAATTATGTAAAACACAT
GCACAAAAAATTTGAAAGTTATTTTGGTAAAATTATAGATAAAATAATCAAAGAGGATAG
TTATAATAATGATGATTTTTTAACCGATAAGAAAAAAGCACTAATATATTCAGTTTTAAT
TTATGAGAATTTTAATAGGGGACTAGTTTATAGATATTTTGAAAAAAATTATTTTGTACT
GGTAGAATAAAAACATTTGGAATAATGCAAGTAACCTCAGCAGAGTACCTTTCCAATGAG
GAAAGTATAAAAAAAGGCGGAAATATTCTTATGGAAAAATACAATGAAAAATATAATGAA
TCTATTGATGGCAATAAAACTCTCTATAAATCATATTATGAATCAAGAAGAGAGAGTATT
AAAAACTACAACCCAGATGCAAAATACATTAATGAAATTGAATCAATTTACATGATGCTT
GGAGAAATCTATCCAAATGCACCAGACTTCATGTCACCACATTTTGAGGGGGACTGCTCT
GAGGGGGAATAAAATCATTATTTATTCTTTATTAGTTATTAGCAGGATTTGTCGGGCATA
AATGCCCGACCTACAAATTCAATTTTTTCAAACCTCTGCCAAATATTTTCATCTTTGCAA
GGCTGTCTGAAAACCCAAACCCCATTTTCAGACGGCCTTTTTTCGCTAAAATCCCCATAC
CGTTCAATCCGAAAACACAGGAGAATCATCATGGAAGTTACCATCTCCGCCATCATCAAT
GGCGAATTTGCCGACCAATACGGCAAGCGCGGTAGTCAGTTTAATGAAAACGGGATGCTG

ATTTAATTCTATTTCCTTTGAAACTACCAATAACCTGCCTCCATCATAAAACTAAAAGCA
AGCCGTAGCCTGCATTCCCACAAACCGCGTGCGTTGCCATGTCACACACCCTACCTGCGG
GCGACGCAAACCTTAAGAGACCTTTGCAAAATTCCCCAAAATCCCCTAAATTCCCACCAA
GACATTTAGGGGATTTCTCATGAGCACCTTCTTTCAACAAACCGCCCAAGCCATGATTGC
CAAACACATCGACCGCTTCCCGCTATTGAAGTTGGACCGGGTGATTGATTGGCAGCTGAT
CGAACAATACCTGAACCGTCAAAAAACCCGTTACCTTAGAGACCACCGCGGCCGTCCTGC
CTATCCCCTGCTGTCCATGTTCAAAGCCGTCCTGCTCGGACAATGGCACAGCCTCTCCGA
TCCCGAACTCGAACACAGCCTCATTACCCGCATCGATTTCAACCTGTTTTGCCGTTTTGA
CGAACTGAGCATCCCCGATTACAGCACCTTATGCCGCTACCGCAACCGGCTGGCGCAAGA
CAATACCCTGTCTGAACTGTTGGAACTGATTAACCGCCAACTGACCGAAAAAGGTTTAAA
AATAGAGAAAGCATCCGCTGCCGTCGTTGACGCCACCATTATTCAGACCGCCGGCAGCAA
ACAGCGTCAGGCCATAGAAGTTGACGAAGAAGGACAAATCAGCGGTCAAACCACACCGAG
TAAGGACAGCGATGCCCGTTGGATAAAGAAAAACGGCCTCTACAAACTCGGTTACAAACA
ACATACCCGTACCGATGCAGAAGGCTATATCGAGAAACTGCACATTACCCCCGCCAATGC
CCATGAGTGCAAACACCTGTCGCCGTTGTTGGAAGGTCTGCCCAAAGGTACGACCGTCTA
TGCCGACAAAGGCTATGACAGTGCGGAAAACCGGCAACATCTGGAAGAACATCAGTTGCA
GGACGGCATTATGCGCAAAGCCTGCCGCAACCGCCCGCTGTCGGAAGTGCAAACCAAGCG
TAACCGATATTTGTCGAAGACCCGTTATGTGGTCGAACAAAGCTTCGGTACGCTGCACCG
TAAATTCCGCTATGCCCGGGCAGCCTATTTCGGACTGATTAAAGTGAGTGCGCAAAGCCA
TCTGAAGGCGATGTGTTTGAACCTGTTGAAAGCCGCCAACAGGCTAAGTGCGCCCGCTGC
CGCCTAAAAGGCAGCCCGGATGCCTGATTATCGGGTGTCCGGGGAGGATTAAGGGGGTGT
TTGGGTAAAATTAGGCGGTATTTGGGGCGAAAACAGCCGAAAACCTGTGTTGGGATTTCG
GTTGTCGTGAGGGAAAGGAATTTTGCAAAGGTCTCCAGCAGTTTGCGCATACATGCCGTA
ACGGCAACCTTATACGGCTTACCCTCGGACAGCGGGCGTTGGTGGAAATCCCGAATAAGC
GGTTCAAAACGTGTCGCTGCCACGGTAGCCATATACAGTGCCTTAAGCACCGCAGACCTT
CCGCCAAAGCAGCGGCTTTTGAATTTGGCTTCCCCGCTCTTCCTCGGGTGCGGGGCAATG
CCGACCAAACTCGCTATCCGTTTGTGCGACAGCCGCCCCAATTCAGGTAGCATCGCCATC
AGCGTAGCCGTCGTTATCGAACCGATGCCTTTGATTTGCTCCGCCACTTGGGCTTTGCCG
TCAAAATGCGTGTGGGTGTGGTCGTCGATTTGTTTGTCCGATTCGTCAATCAGCCGGTCA
AAATGGGCAATCAGTTGTTTGACGCTTCCGACTTGCGTTTCGTGAACCTGATGCAGACGG
TTTTTCTCGGCAGTCCGCATATCCGCCGATTGGTTGCGGCGGTTAACCAAGGCTTCCAAC
ACTTCTTCCGCTTCTGTGGGCGGGTGGTAGGGCATGGTTTGCCAATCTTCTTTCTGTGCC
TTCATCTGTGCGAAGAAGGCAGGCATTTTGGCATCTTTGGCGTCGGTTTTGGTCAGCGAC
TGCGATTGGGCAAACTGATGCGTCTGACGCGGGTTGGCGATAATCACGGCTATGCCTGCT
CGGTGGATGGCTTTGGCGGCGGGGATTTCGAGACCTCCGGTACTTTCCGTCACGACGAGG
GCGACCTTGTGTTTTTTAAGGTATTCGATAGTATGGGCGATACCTTTGGGGTTGTTGGTT
TCGGTTTTGGTTTTAGACAAAGACGAAACGGCGATGACGAAGTTTCGTTTGGCGATGTCG
ATATAGTGAATTAACAAAAATCAGGACAAGGCGGCGAGCCGCAGACAGTACGGATAGTAC
GGAACCGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTCGAGCTAAGGCGA
GGCAACGTCGTACTGGTTTTTGTTAATTCACTATATCTGTGCGTTACGACGGCATGCCGT
CTGAAGGGTGTTTATGTCTGCATCTAAGAAATTTCCGATTCCTTTGAGCTATTTCAGCAT
CGCGCTGGGCTTGTTTGCCTTGGGGCTGTCGTGGCGTTACGGCGCGTCTGTCGGGCTGCT
GCCCGCCTTGGCCGCCGAATCGCTGCTTGCGGCGGCTTCGGTCGTCTGGCTCTTGCTGGT
GGCGGCATACCTGATCAAAATGTTTGCGTACCGAAACGATTTTTTGTCTGATTTACGCGA
CTTGGTGCAATGCTGCTTCATCAGCGCGATTCCGATTACCGCTATGCTGGAGGGACTCGC
GCTGAAGCCCTATCAGGCAGGCGCGGCGGCAGTCCTGATTTATGTCGGCGTTGCCGGACA
GTTGGCTTTTTCGATGTATCGGGCGGCCGGTCTGTGGCGCGGCCTGCATTCCTTGGAGGC
GACGACGCCGATTATTTATCTGCCTACGGTTGCGACAAACTTTGTCAGCGCGTCATCTCT
GGCGGCGTTGGGGCATCATGATTATGCAGCTTTGTTTTTCGGCGCGGGTATGTTTTCCTG
GCTGAGCTTGGAAGCCTCCATCTTGGGCAGGCTGCGCACGGCGGCACCGGTCGGCACGGC
CGCGCGCGGCGTGGTCGGCATCCAGCTTGCGCCCGCCTTTGTCGGCTGCGGCGCGTATTT
TGCCGTCGGCGGTAAAGTCGACGGTTTTGCGTTGGCATTAATCGGCTACGGCTGCCTGCA
GCTTTTGTTCTTGCTGCGCCTGACCCGCTGGTTTTGGGAAGGTGGTTTTACGATGAGCTT
TTGGGGATTTTCATTCGGTTTCGCGGCAATGGCAGGATGCGGTCTGCATCTGGCGGCTTC
CGGCGTATTGTCGGGCTTGGGGCTGACGCTTGCCACCGCCGGATCGGCAGGCGTGGCGCT
GCTGCTTGTCGGTACGCTGCACCGGATAGCGACGGGGCGTTTCTTGGTACGCAGCTGATG
CGTTTTGCCGCCTTGTCAAAAATGCCGTCTGAAACGCTGGGATTCAGACGGCATTTTTTA

TTTCACACCCTTACAGGTAGAATTTTTCGATGACTTTCAAATTGTCGTCCAATTTGTACA
CCAACGGCTGACCGGTCGGGATTTCCAAGCCCATAATGTCTTCGTCGGAAATGCCCTCGA
TGTGTTTTGCCAGCGCGCGCAGGGAGTTGCCGTGCCCCGCCACCAAGACGCGTTTGCCGC
TCAAAATCGCGGGGGCGATTTGGTCTTCCCAAAACGGCAATACGCGCTCCAGCGTTACTT
TCAGGTTTTCGCCGTCGGGTACGACATCGGCAGGCAGATGGGCATAGCGGCGGTCTTTGT
GTGCGGAAAACTCATCGTCTTTGTCCAAAAGCGGCGGCAGGGTGTCGTAGCTGCGCCGCC
AGATCCGGACTTGCTCGTCGCCGTATTGTTCGGCGGTTTGTTTTTTGTCCAGGCCTTGCA
GTTGGCCGTAGTGGCGTTCGTTCAGCCGCCACGTTTTGATTTGCGGTACGAACAGTTGGT
CGGATTCTTCCAAAACGATGTTGCAGGTCTTAATCGCGCGGGTCAGGACGGATGTGAAGG
CGATGTCGAACTCATAGCCGTTTTCTTTCAGTTTCTTGCCGGCGGCGGCAGCCTCGGCAA
GCCCCTGCTCGCTCAGCTTCACGTCGCGCCAGCCTGTAAACAGGTTTTTCGCGTTCCATT
CGCTTTGTCCGTGGCGGATAAATACCAGTTCCATATCGTCTCCAATGTGTGAAAGTGGGA
AAGCCTTATTTATAACATATTTTCACATTTCCCGTATTTGATTCAGATTCAGACACGCGC
CCACTATGGTTTGCCGTTTTGATTTACAATAATGTCCTTTGCTTTACATTCCGCATACAC
AATGAATACGCAAGCGCACGCCCCACATACCGATTCCAATACGCTGATGCTCGGCCGATA
CGCCGAACGCGCCTATCTCGAATACGCCATGAGCGTGGTCAAAGGCCGCGCGCTGCCTGA
AGTTTCAGACGGCCAGAAGCCCGTGCAGCGGCGCATTTTGTTTGCCATGCGCGATATGGG
TTTGACGGCGGGGGCGAAGCCGGTGAAATCGGCGCGCGTGGTCGGCGAGATTTTGGGTAA
ATACCACCCGCACGGCGACAGTTCCGCCTATGAGGCGATGGTGCGGATGGCGCAGGATTT
TACCTTGCGCTATCCCTTAATCGACGGCATCGGCAACTTCGGCTCGCGCGACGGCGACGG
GGCGGCGGCGATGCGTTACACCGAAGCGCGGCTGACGCCGATTGCGGAATTGCTGTTGTC
CGAAATCAATCAGGGGACGGTGGATTTTGTGCCGAACTACGACGGCGCGTTTGACGAACC
GCTGCACCTGCCCGCCCGCCTGCCTATGGTGTTGCTCAACGGCGCGTCAGGCATTGCGGT
GGGCATGGCGACCGAGATTCCGCCGCACAATTTGAACGAAGTGACGCAGGCGGCGATTGC
GTTGTTGAAAAAGCCGACGCTGGAAACCGCCGACCTGATGCAATATATTCCTGCCCCCGA
TTTTGCCGGCGGCGGTCAAATCATCACGCCGGCGGACGAATTGCGCCGGATTTATGAAAC
CGGCAAGGGCAGCGTGCGCGTGCGTGCGCGTTATGAAATCGAAAAATTGGCGCGCGGACA
GTGGCGCGTCATCGTAACCGAGCTGCCGCCGAACGCCAATTCCGCCAAAATCCTTGCCGA
AATCGAAGAGCAAACCAACCCGAAACCGAAAGCGGGTAAGAAACAGCTCAACCAAGACCA
GCTCAATACCAAAAAGCTGATGCTGGATTTAATCGACCGCGTGCGCGACGAGTCCGACGG
CGAACATCCCGTGCGACTGGTATTCGAGCCGAAATCCAGCCGCATCGATACCGATACCTT
CATCAACACGCTGATGGCGCAAACTTCGCTGGAAGGCAATGTGTCGATGAACTTGGTGAT
GATGGGTTTGGACAACCGCCCCGCGCAGAAAAACCTGAAAACGATTTTGCAGGAATGGCT
GGATTTCCGCACCGTAACCGTAACACGCCGTCTGAAATTCCGTTTGAACCAAGTGGAAAA
ACGGCTGCACATCCTCGAAGGCCGTCTGAAAGTCTTTCTGCACATCGACGAAGTGATTAA
AGTCATCCGCGAATCAGACGACCCGAAAGCCGATTTGATGGCGGCGTTCGGGCTGACCGA
AATCCAAGCCGAAGACATTTTGGAAATCCGCCTGCGCCAGTTGGCGCGTTTGGAGGGTTT
CAAACTCGAAAAAGAATTGAACGAGTTGCGCGAGGAACAAGGCCGTCTGAACATCCTTTT
GAGCGACGAAAACGAAAAACGCAAGCTGATTGTCAAAGAGATGCAGGCGGATATGAAACA
ATACGGCGACGCGCGACGCACGCTGGTGGAAGAGGCCGGACGCGCCGTGCTGACGCAGAC
CACCGCCGACGAACCCATCACGCTGATCCTGTCGGAAAAAGGCTGGATACGCAGCCGCGC
CGGACACAATCTCGATTTGAGCCAAACCGCGTTCAAAGAAGGCGACTGCCTCAAACAAAC
CCTCGAAGGCAGAACGGTTTTACCCGTCGTCATCCTCGATTCATCGGGCAGAACCTACAC
GCTCGATGCCGCCGAAATCCCCGGAGGGCGCGGCGACGGCGTACCGGTTTCCTCCTTAAT
CGAGCTGCAAAACGGCGCGAAACCCGTTGCGATGTTGACAGGATTGCCGGAACAACATTA
TTTATTATCAAGCAGCAGCGGCTATGGCTTCATCACCAAGCTGGGCGATATGGTCGGGCG
CGTGAAAGCGGGCAAAGTGGTGATGACCGCAGACAGCGGCGAAACCGTTTTGCCGCCGGT
TGCCGTCTATGCCTCCTCGTTCATCAACCCCGACTGCAAAATCATTGCCGCCACCAGTCA
AAACCGCGCCCTCGCCTTCCCCATCGGCGAATTGAAAATTATGGCGAAAGGCAAAGGGCT
GCAAATCATCGGATTAAACGCCGGCGAATCGATGACGCATACCGCCGTTTCTTCCGAGCT
GGAAATCCTGATTGAAAGCGAAGGCAGGCGCGGCGCGGCGCACAAAGACCGCATCCCCAT
CTCCCTGCTTGAGGCAAAACGCGGCAAAAAAGGCAGACTATTGCCCATATCGGGCAGCCT
GAAACAGCTTTCTTCCCCTAAATAAACCCGGTTCCGCACATATTATGGTGATTTCCAACC
CCCGCGAACTTGAAAAACTCAAAGACCGGATTCCCAATCTGATCAACATCATCCGCGTCG
CCATCGTTTTTCCGCTGATGATTATGCACATCCTCGGGCTGGAAACCGGCAGCCGTGCGA
ACCTCCACGCTTCGTGGACGGCGTGGGCGTTTTATGTTTGGCTCGCCATTGCCTGCTGGC
TGATTTTCTTTTCCATTATCCATCCGCATTGGCAATGGCAGTCGCTGAAAATGCCGCGTT

TCAGCGCGGTAGCGGACATCACGATGATCGGCGTGCTGACCTACCTGTTCGGCGGCATCG
ATTCCGGCTTCGGCATCCTGATCCTGCCCTTCGTCGTCTGCTCCTGCCTGCTCAGCTACG
GGCGCTACCCCCTGCTCTATTCCAGCTACGCCGCCATCCTGCTGATATTCAACGCCATTG
CCGACGGCGATATCGGCAAATACCCGCTCATATCGGATGCCCGAACCGCCTCGGCAACCT
TCATCCTTGTCGCCGCCTCCTATCTTTCCGCCATCTTCACCTCACTGTCGGTCAAATACA
TCGACCGTGCCGGAAAACTCGCCTACGACAGCCATATCGCCTACCACCGCATCAAAGGCT
TGAGCCAAACCGTACTCGAACGCGTTCAGGAAGCTGTCGTCGTCATCAATGCCGAAGGGC
TGGCGGTGCTGTTCAACCGGAAGGCGAAAGACCTTTTCCCCGCGCTCGAAATCGGACGGC
GCGCCGGTCTGTCCGATTCTGCCGCCGAACTGTGGGATCAAGCCTCTCCGCACACTTTCG
AATACGTCCTCGGCACACCCGGCCTGAACGCCGGCATCCGCGCCGTTCCGGTCAACAAAG
GGTCGGACAAGCTGCTCATCCTCTACATCCGCCCGCAAAGCGAAATTCAGGCAGAAGCCC
TGTCCGTCAAACTTGCCGCGCTCGGACAACTGACCGCCAACCTCGCCCACGAAATCCGCA
ACCCGATGTCCGCCATCCGCCACGCCAACGACCTGCTGCGCGAAAATATGGAAGCGGGGG
CGGCAGATCCGTTCAACGCCAAATTGTGCAAAATCATCGACGGCAACATCTGCCGCATCG
ACAAAATGCTCGAAGACATTTCCTCGCTCAACAAGCGCAACAAAACCGAACGCGAAACCA
TCGGCCTGATACCGTTTTGGGAAGAATTCAAACAAGAGTTCCTGCTCGGCCATCCCGATG
CCGCCGACTGCATCCGTCCGGACATTCAAGGCGGCAGCCCGACCGCCTATTTCGATCCCG
CCCACCTGCGGCAAATTATGTGGAACCTCGCCAACAACGCGTGGCGGCACAGCCGCAAAC
AGCCCGGCTCGATTTCCGTCACCATCCGCCCCGCGCAAAAAAACACCGTCTGTATCCTCT
TTGCCGACCGCCCGAAGTGCAGGAACACCTGTTCGAACCCTTTTACACCACGGCGGAAAA
CGGCACCGGCCTCGGGCTGTATGTCGCCCGCGAACTGGCGCACGCCAATTTCGGCGATTT
GACCTACCTACCGGAAGCCAAATGTTTCGAACTCACATTACCGGAAAAAACCAATGACTG
AACTGCAACACCCCGTCCTCGTCGTCGATGACGAAACCGACATTCTCGACCTGATGGAAA
TGACCCTGATGAAAATGGGCTTGCGCGTCCATACCGCGTCAGGCGTTGCCGAAGCCAAAA
ACAAGCTCGACAGCCAACGCTATTCGCTCGTCCTGACCGATATGCGTATGCCGGACGGCT
CGGGGCTGGAAGTCGTCCAACACATCAACAGCCGCCTGCTCGATACGCCGGTTGCCGTCA
TCACCGCCTTCGGCAACGCCGATCAGGCACAGGAAGCGTTGCGTTGCGGCGCGTTCGACC
CCGATACCATGCAGATACAGGACTATCTCGACCAAATCGAACGCGACATCATCGAACAAA
CCCTCAAACAAACCGAAGGCAACCGCACGCAGGCCGCCAAACGCTTGGGCATCAGCTTCC
GTTCCATGCGCTACCGTATGGAACGCCTCAACATCGGCTGACGACAAAACGGCATCCGCA
CCATCTCCGCCCACCCGAAAAAATGCCGTCTGAAACGGCACGGGAAAGCGGGTTCGCCCC
ACGCCCGAACGGACACAAAACACCATGACCGACATCCTTATTGACAACACCGCCACCGAA
ACCGTCCGCACCCTGATACGGGCATTCCCCCTTGTGCCCGTTTCCCAACCGCCCGAACAA
GGCAGTTACCTCCTTGCCGAACACGATACCGTCAGCCTCAGGCTTGTCGGGGAAAAAAGC
AGCGTCATCGTCGATTTTGCCTCCGGCGCGGCACAATACCGGCGCACAAAAGGCGGGGGC
GAACTCATCGCCAAAGCCGTCAACCACACCGCGCACCCCACCGTTTGGGACGCAACCGCA
GGATTGGGGCGCGACAGCTTCGTCCTCGCCTCGCTCGGGCTGGCCGTTACCGCCTTCGAG
CAACATCCCGCCGTCGCCTGCCTGCTTTCAGACGGCATCCGCCGCGCCCTCCTCAATCCC
GAAACGCAAAACACCGCCGCGCACATCAACCTCCATTTCGGCAACGCCGCCGAACAAATG
CCCGCACTTGTCCAAACACAAGGCAAACCCGACATCGTCTATCTCGACCCCATGTATCCC
GAACGCCGCAAAAGTGCCGCCGTTAAAAAAGAAATGACCTACTTCCACCGGCTCGTCGGC
GAAGCGCAAGATGAAGCGGCACTCCTGCATACCGCACGCCAAACAGCAAAAAAACGCGTC
GTCGTCAAACGCCCCCGCCTCGGCGAACACCTTGCCGGACAAGACCCTGCCTACCAATAC
ACAGGCAAAAGCACCCGCTTCGACGTTTACCTGCCCTACGGGACGGACAAGGGATAACGC
CCATAAAACAAGACACCGAAAAATTTGCCGTTCTTATGCAAACGAGAAACCGGTTTTTGC
GTTTCGACTGTTTTGGATAAGTCATCACACCTTAAAGTTTGTCATTCCCACAGAAGTGGG
AATCCGATTCATTCAGTTTTATAGTGGTTTAAATTTAAACCACTATAGTTGTTTTCGAGT
TTCAGGCAACTTCCAAACCGTCATTCCCACGGAAGTGGGAATCTAGAAATGAAAGGCAAC
AGGAATTTATCGTAAATGACTGAAACCGAACGGACTAGATTCCCGCCTACGCGGGAATGA
CGGGGCGGGCAGATGCCGTCTGAAATTCCGTCATTCCCGTGAAAACGGGAATCTAGAACT
TCTGATTTTTCAGACGACTTTTGAACATTGCCGCCACCCAATGATCTGGATTCCCACCTG
CGCGGGAATGACGAGGTTTCAGGTTGCTGTTTTTAAGTTGCTGTTTCGGGTTGCTGTTTT
TTATGGAAATGACAAGGTTTTAGATTGCGAGAATTTATCCGCTCCTCCGTCATTCCCACG
GAAGTGGGAATCCAGAAATGAAAAGCAACAGGAATTTATCATAAATGACCGAAACCGAAC
GGACTAGATTTCCGACTGCGCGGGAATGACGGGGCGGGAGGATGCCGTCTGAAATTCCGT
CATTCCCGTGAAAACGGGAATCTAGAACTTCTGATTTTTCAGACGACTTTTGAACATTGC
CGCTACCCAATGATTTGGATTCCCGCCTGCGCGGGAATGACGATGTAAAATTATCCGGGA

TTCAAAAAGACAGGCTTTCACATCCGTGGGAATGACTGCGGAAAGATGATTTTTATAGTG
GATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAAATAGTACGGCAAGG
CGAGGCAACGCCGTACTGGTTTTTGTTAATCCACTATATTTGTCATAAAAATCCGCACC
TTAATCAGTTGGCGGTTAAATCAAACTTTTAGGGTGCAGATTACTTTTTATGATTTCAGA
CAGCATTTTGACAGGCGGCAGCCTATTTCGGCAATACCAAAAACTTAATCAGCAGTTCTT
TGAATACAAAACCGAACACGCCCAAGCCCAAAACCAAAAACAAAATGGCGATGCCGAATT
TGCCTGCTTTGGACTCCTTGCCCAAATTCCAAACGATAAAACCCAAAAAAATAATCAAGC
CGGTCAGGCAGATTTTCAACGCCCAATCGGCAAAAACCGCTTCATCCATATTTTTTTCCT
ATTGTTGATGTGTATGCCATATAAGATAAGGGTTTCAGACGGCATCTGCTGTCCAATGCC
GTCTGAAACACGCAATCAGCGTGCGAGTGCCTGTTTCAAATCGTCAATCAAATCGCCAAC
ATATTCCAAACCGACCGACAGGCGCACCAATCCGGGGCGGATGTTGGCGGCGAGTTTTTC
TTCGGGCTGCATCCTGCCGTGCGTGGTTGTCCACGGGTGGGTAATGGTCGAGCGCACGTC
ACCGAGGTTGGCGGTGCGGGAAAAGAGTTCCACGCCGTCCACAACTTTCCACGCCGCTTC
TTGATCGGCAACTTCAAAGCCGATGACGATGCCGCCGCCGTTTTGCTGTTTGCGGATAAG
CGCCGCCTGAGGATGGTCGGACAATCCGGTGTAGTACACGGCTTGAACCTGCGGCTGCGC
TTGCAGCCATTGTGCGATTTTCAGGGCGTTGTCGAACTGTTTTTCCATACGCAGCGACAG
GGTTTCCACGCCGCTCAACAACTGCCACGCATTAAACGGCGACATCGCCAGCCCGCAAGA
GTTGCAATACATGGCGACCTGCGCCAACAACTCTTCCGAACCCGCCAACACGCCGCCCAT
CACACGCCCGTGTCCGTCTATGGCTTTGGTCGCGGAGGAAACGGAAATATCCGCACCGTG
TTTCAAAGGCTGCGAGCCGACGGGCGACAGCAGGCTGTTGTCCACCACCAAGAGCGCGCC
GATGCCGTGCGCCAATTCCGCCAAGGCTTCCAAGTCGGCCACTTCGCCTAAGGGGTTGGA
CGGCGTTTCCAAAAACAGCAGTTTGGTATTGGCTTTGACGGCGGCTTTCCATTCGTTTAT
ATCAGTCGGCGACACGTGGCTCACTTCGATGCCGAATTTGGCAACGATGTTATTGATAAA
GCCGACGGTCGTGCCGAACAGGCTGCGGCTGGAAATCACATGGTCGCCCGCCTGCAAAAA
GGTGAAAAACGCCGCCTGAATCGCAGACATACCCGCCGAAGTGGCGACCGCGCGTTCCGC
ACCTTCCAAAGCGGCGATGCGTTTTTCAAAGGCGGCTGTGGTCGGGTTGGCGGTACGGGT
ATAAGTGAACCCTTTGATTTTTTTTGAAAACAAATCGGCAGCGTGTTGGGCGTTGTCCCA
CATGAAGCTGCTGGTCAGAAACAATGCCTGATTGTGTTCGCGGTATTCGGTTTGTTCTTT
GCCGCCGCGTATGGCGAGCGTTTGCGGATGGAGTTTTTGCTCATCGGTGATTCCTCGGT
TTTGTCCGTTCGGCAACGGAGCGTGCGCCCGTTGTTTAATTTGTTAATATTTTGCGCCTG
TTCTATGATGCTTTCAAGTCGGATGAGAATGCAAATGCCGTCTGAAACGGCTTTCAGACG
GCATGGCAATCAGCGTTTGTATTTTAACTCGTACTTGATGTCGTTGAGGATTTTGCGGAC
ATCGTGTTCCAACACGTCTTCGACTACCGCCCCCGCCTGCTCGTGCAGCATCTGCTGGAG
CTGATAGGTGAAAACCGCCATCTGCTTTTGCACCGCCGTTCGGATGATGCCGTTGACGGT
ATCGGTCAGATGCGGGCGCAGGCGTTTGATCAGCCGTTCGGTCAGCTCCTGTTCGGACAG
GCAGAACACTTCGCGCCGGTTGACGGCTTTCGGGTTCAGGATATTGATTTGGACGGGCAT
CAACGTTTCTTCCGCATCGTTTTCCCCGTTTTCCGAAACCGCCGGCTCATTCGTGCCGGA
TTCTGCCTCGTCGGCGTTTTCCCCGCTTTCAATCTGTCCGGTTTCAAATTCGACACTGTC
TTTTTTGGTATCAAACCGGATTCTCCGCCGCGATTCGATGTGTTTTTCCGAAACCGACAT
TTGCAGGGAAGCCTGCGCGTTGAGCCAGTTTTCCTGAAGGACGATCATCGGGTCGGTTTC
GACTTCCTCGCCGCAATCGGCAACGGCGGCATTGTGTTCCTCCTGCCATTTTTTCAGATA
CGCCTTCAACACACGGGCTCGGCTCTCATCGTCCAGTTTCGGCACAGGCGCGTCCGTTCC
GGTTTCAGAGGGGCGGGACAGCGGCGCGTAAGTCGGCACTGCCTTCATACGGCGCGTCTG
ACGCAGGTTTTCCAAACGTTTTTCCCAATTCGGCTCTTTATTCGCATCCATTTTCGGCTT
CCGGTTCTTAATCTTTGCAAGCAGACAAACCCGCGCCCAAAGCGCGGTTTGATATAATGG
CGCATTTTAACAGATTCGCGAGGATACATCATGGGCAGCATCGAACAGCGTTTGGAATAT
CTGGAAGAGGCGAACGACGTGCTGCGTATGCAGAACCACGTCCTGTCCACCGCATTCAAA
GCCTTAATCCGCGCCCTTCCCGCCGAAACCGCCGAAATCGCGGTCGAGTCGATTCAGCTT
GCTTTTGAGGACGCCTTGGCAGAATTGAGCTATGAGGACAGCCCGCATACGGATTTGTTC
CACGACGTTACTTATGCGTTTTTCCGTGAAAAAGAACGTTAATTTTATGTTAAACTGATT
TTTTAGGCTTTTTGATTACCGAAAGGAATTTTGATGAATATGAAAAAATGGATTGCCGCC
GCCCTTGCCTGTTCCGCGCTCGCGCTGTCTGCCTGCGGCGGTCAGGGCAAAGATACCGCC
GCGCCTGCCGCCAACCCCGACAAAGTGTACCGCGTGGCTTCCAACGCCGAGTTTGCCCCC
TTTGAATCTTTAGACTCGAAAGGCAATGTCGAAGGTTTCGATGTGGATTTGATGAACGCG
ATGGCGAAGGCGGGCAATTTTAAAATCGAATTCAAACACCAGCCGTGGGACAGCCTTTTC
CCCGCCTTAAACAACGGCGATGCGGACGTTGTGATGTCGGGCGTAACCATTACCGACGAC
CGCAAACAGTCTATGGACTTCAGCGACCCGTATTTTGAAATCACCCAAGTCGTCCTCGTT

CCGAAAGGCAAAAAAGTATCTTCTTCCGAAGATTTGAAAAACATGAACAAAGTCGGCGTG
GTAACCGGCTACACGGGCGATTTCTCCGTATCCAAACTCTTGGGCAACGACAATCCGAAA
ATCGCGCGCTTTGAAAACGTTCCCCTGATTATCAAAGAACTGGAAAACGGCGGCTTGGAT
TCCGTGGTCAGCGACAGCGCGGTCATCGCCAATTATGTGAAAAACAATCCGGCCAAAGGG
ATGGACTTCGTTACCCTGCCCGACTTCACCACCGAACACTACGGCATCGCGGTACGCAAA
GGCGACGAAGCAACCGTCAAAATGCTGAACGATGCGTTGGAAAAAGTACGCGAAAGCGGC
GAATACGACAAGATTTACGCCAAATATTTTGCAAAAGAAGACGGACAGGCCGCAAAATAA
GCCCGCCCGTCCGAACACAATGCCGTCTGAAGCCCTTTCAGACGGCATTGTTCATCAATC
GGCCTACAATGAACTGCCTGCTGATTTCTCCCTACCGCAAAGCAACAGGCAAAGATTACA
AATATCAAAATCCGAGTAAAACAGTATTTTATTAAAACAAATTGATAATCAAGAGATTAG
AATTATGTATTGTCTTTACCGTACAAACGCTGGCACTATTTCAACCTGATAAAAAACAGC
CTTCAAAAAGGTTGTTTAAAACAGCAGCAGACACTTACCGCCACAACCTTGAAAAGGAAC
ACAATCATGACCGTCATCAAACAGGAAGACTTTATCCAAAGCATTTGCGATGCCTTCCAA
TTCATCAGCTACTATCATCCCAAAGACTACATCGACGCGCTTTATAAGGCGTGGCAGAAG
GAAGAAAATCCTGCCGCCAAAGACGCGATGACGCAGATTTTGGTCAACAGCCGTATGTGT
GCGGAAAACAACCGCCCCATCTGCCAAGACACAGGTATCGCAACCGTCTTCCTCAAAGTC
GGTATGAACGTCCAATGGGATGCGGACATGAGCGTGGAAGAGATGGTTAACGAAGGCGTA
CGCCGCGCCTACACTTGGGAAGGCAATACGCTGCGCGCTTCCGTCCTCGCCGATCCGGCC
GGCAAACGCCAAAACACCAAAGACAACACCCCCGCCGTCATCCATATGAGCATCGTGCCG
GGCGGTAAAGTCGAAGTAACCTGCGCGGCAAAAGGCGGCGGCTCTGAAAACAAATCCAAA
CTCGCCATGCTCAATCCTTCCGACAACATCGTCGATTGGGTATTGAAAACCATCCCGACC
ATGGGCGCGGGCTGGTGTCCTCCCGGCATCTTGGGTATCGGCATCGGCGGCACGCCCGAA
AAAGCCGTGCTGATGGCAAAAGAGTCCCTGATGAGCCACATCGACATTCAAGAATTGCAG
GAAAAGGCCGCGTCCGGCGCGGAATTGTCCACCACCGAAGCCCTGCGCCTCGAACTCTTT
GAAAAAGTCAACGCGCTGGGCATCGGCGCACAAGGCTTGGGCGGACTGACCACCGTGTTG
GACGTGAAAATCCTCGATTATCCGACCCACGCCGCCTCCAAACCGATTGCCATGATTCCG
AACTGCGCCGCCACCCGCCACGTCGAATTTGAATTGGACGGCTCAGGCCCTGTCGAACTC
ACGCCGCCGCGCGTCGAAGACTGGCCCGATTTGACTTACAGCCCCGACAACGGCAAACGC
GTCGATGTCGACAAGCTGACCAAAGAAGAAGTGGCAAGCTGGAAAACCGGCGACGTATTG
CTGTTGAACGGCAAAATCCTCACCGGCCGCGATGCCGCACACAAACGCCTCGTCGATATG
CTCAACAAAGGCGAAGAATTGCCCGTCGATTTCACCAACCGCCTGATTTACTACGTCGGC
CCCGTCGATCCGGTCGGCGATGAAGTCGTCGGTCCGGCAGGTCCGACCACAGCCACCCGC
ATGGACAAATTCACCCGCCAAATGCTCGAACAAACCGACCTCTTGGGCATGATCGGCAAA
TCCGAGCGCGGCGTGGCCACCTGCGAAGCCATCGCCGACAACAAAGCCGTGTACCTCATG
GCAGTCGGCGGCGCGGCGTATCTCGTGGCAAAAGCCATCAAATCTTCCAAAGTCTTGGCG
TTCCCCGAATTGGGCATGGAAGCCATTTACGAATTTGAAGTCAAAGACATGCCCGTAACC
GTCGCCGTAGATAGCAAAGGCGAATCCATCCACGCCACCGCCCCGCGCAAATGGCAGGCG
AAAATCGGCATCATCCCCGTCGAATCTTGAGGCGCCATGCCGTCTGAACACAAAATCTGC
CTTCAGACGGCATTTCCGCCCCCGGTTGCGGTACAATCCACCATTTCATCACTCGGCGAC
CCACACCGTGAAAATCCTCATTTTAGGCAACGGACAGGTAGGTTCTACCGTCGCACAAAA
CCTTGCCGCCATACCCAACAACGACGTAACCGTTATCGACATCGACGAAAAAGCATTGCA
GGAAACAGGCAGCCGCCTCGATGTCCAAACCGTTTTCGGCAACGGCGCATCCCCCTTCAC
ATTAGAACGCGCCGGCGCGGAAGATGCCGACTTGCTGCTCGCGCTCTCCCGCAGCGACGA
AACCAACATCGTCGCCTGCAAAGTTGCCGCCGACCTGTTCAACATCCCCGGCCGCATCGC
GCGCGTCCGTTCCAGCGAATACCTCGAATACCTCAGCCCCAAGCTCGAAAACAACGAAAA
CGGCAGCCTTTCCATATTCGGCATAACCGAAACCATCAGCCCCGAACAGCTCGTTACCGA
ACAGCTTGCCGGCCTGATAGACTGCCCGGGCGCATTGCAGGTTTTACGTTTTGCAGACGA
CCGCGTGCGGATGGTCATCATACAGGCGCGGCGCGGCGGACTGCTTGTCGGACGCAGCAT
TGCCGACATCGCCCAAGATTTGCCCGACGGGGCCGACTGCCAAATCTGCGCCGTTTACCG
CAACAACCGCCTCATCGTCCCCGCGCCGCAAACCGTCATCATCGAAGGCGACGAAATCCT
ATTTGCCGCCGCCGCCGAAAACATCGGCGCGGTCATACCCGAATTGCGCCCCAAAGAAAC
CAGCACCCGCCGCATCATGATTGCCGGCGGCGGCAACATCGGCTACCGTCTCGCCAAGCA
GCTCGAACACGCATACAACGTCAAAATCATCGAATGCCGGCCGCGCCGTGCCGAATGGAT
AGCCGAAAACCTCGACAACACCCTCGTCCTGCAAGGTTCGGCAACCGACGAAACCCTGCT
CGACAACGAATACATCGACGAAATCGACGTATTCTGCGCCCTGACCAACGACGACGAAAG
CAACATTATGTCCGCCCTTTTGGCGAAAAACCTCGGCGCGAAGCGCGTCATCGGCATCGT
CAACCGCTCAAGCTACGTCGATTTGCTCGAAGGCAACAAAATCGACATCGTCGTCTCCCC

CCACCTCATCACCATCGGCTCGATACTCGCCCACATCCGGCGCGGCGACATCGTTGCCGT
CCACCCCATCCGGCGCGGCACGGCGGAAGCCATCGAAGTCGTCGCACACGGCGACAAAAA
AACTTCCGCCATCATCGGCAGGCGCATCAGCGGCATCAAATGGCCCGAAGGCTGCCACAT
TGCCGCCGTCGTCCGCGCCGGAACCGGCGAAACCATTATGGGACACCATACCGAAACCGT
CATCCAAGACGGCGACCACATCATCTTTTTCGTCTCGCGCCGGCGCATCCTGAACGAACT
GGAAAAACTCATCCAGGTCAAAATGGGCTTTTTCGGATAAACCGCCCCATTCCGGACATA
TTGCCGCCAAGCGGTATGGAAGCGGAAATAATGGTAGGTGGGCTTCAGACGGCATCCGCC
CTCCCCGTCATTCCCGCGTAAGCGGGCATCCAGACCTTGGGATAGCGGCAATATTCAAAG
GTTATAAAAGACCCGTCATTCCCGCGCAGGCGGGAATCCAGACCTTGGGATAGCGGCAAT
ATTCAAAGGTTATCTGAAAATTTAGAGGTTCTAGATTCCCGCTTTCGCGGGAATGACGAA
AAGTTGCGGGAATCCAGAACGTCGGGCAACGGCAATATTCAAAAGCCGTCTGAAAATTTA
AAAGTTCTAGATTCCCGCTTTCGCGGGAATGACGAAGTTTCAGACGGCATCGCCCGCCTG
TTTTGATATAGCGGCACCCCCCCGACAAAAAAACAATCCGGAACGCATCTGACCGTTCCG
GCTTGTTTTCAGGCGAATCCGCCGCATCAGAACATACTGCGCACGCCCATATTGACCTGC
CAAGTCTAGCGCATCGTGTGCATCGAAGACCTTTGCGCCTCAAAATAAAGCTGCCTTCCG
TTGTCGGCATTACCACGCAAAAAAATGAATTGCTTGATATTCCAATGTTTTTTATATGTT
TTTATATTGTGATGCGATCAGACAAACGCCCCCCTGACATTTGTTTAGACGGCATCGTAT
TGCTAAATTTCTATAAGTATGTATAATGTCCGTTTCCACGCGCCCATCGTCTAGAGGCCT
AGGACACTGCCCTTTCACGGCGGCAACCGGGGTTCGAATCCCCGTGGGCGTGCCAATTCA
AAAACCTGCTTGTTTCAAGCAGGTTTTTTATTATGAGTCGTCATTCCCGCAATTTTTCGT
CATTCCCGCAAAAGCGGGAATCTAGAGCGTAGGGTTGAAGAAACCGTTTTATCCGATAAG
TTTCCGTGCCGACAGGTCTGGATTmCCGCCTGCGCGGGAAGGACGGCAGAGGGTGGACGA
TGCCGTCTGAAGCCTGACAAAGCATTTGATGCCGTCTGAAACTTCGTCATTCCCGCAAAA
GCGGGAATCTAGAGCGTAGGGTTGAAGAAACCGTTTTATCCGATAAGTTTCCGTGCCGAC
AGGTCTGGATTCCCGCTTTCGTAGGAATGACGGAATTTTAGGTTTCTGTTTTTGTGGAAA
TGACGAATAAAGCGTGCCGGTTTATGCTCGCCGCAACACGCGGTTCAGACGGCATTGCTC
TCTTTTTTCATTATCAGTGGGTGTAGCAACTGTATTTTTCACCCCGTCGGGCAAAAATAC
AGTTGCTACGATGCACCCCGCCGCCCTGCCCTGTGCCTTGTCCTGCAATACGGCATATAA
TGCACCACAAACCCCCGCGCTGCGGTTTTCAGACGGCATCGCCGTGCTTTTTTACAGGCA
TTAGCCCTTTTTATCGGACGCAATATTAAGGAGGAACAAATGAAAAGCTCTTTTGTGCAA
ACGCTTACCATCGCCGGTTCGGATTCGGGCGGCGGTGCGGGCATTCAGGCGGATTTGAAA
ACATTTCAGATGCGCGGCGTGTTCGGAACGTGCGTCATCACCGCCGTTACCGCGCAAAAT
ACCTTGGGCGTGTCGGCGGTTCATCTCGTCCCGACCGAAACCATCACCGCACAAATCCAA
GCAATCAGGGAAGACTTCGACATCCGCGCCTACAAAATCGGTATGCTCGGCACGGCGGAA
ATCATCGAATGCGTTGCCGACAAGCTGAAACACTGCAGCTTTGGCAGGCGCGTACTCGAC
CCTGTGATGATTGCCAAAGGCGGTGCGCCGCTGTTGCAGGATTCCGCCGTTGCGGCACTG
ACGCGCCTGCTGCTTCCCGATACGGATGTATTGACCCCCAACCTGCCCGAAGCGGAAGCT
CTGACCGGCGTGCATATTGAAAACCGTAAAGATGCGGAACGTGCGGCAAAAATCCTGCTT
GATTACGGTGTCAAAAATGTCGTTATCAAAGGCGGACATTTGAACGGCAGCACAAGCGGA
CGCTGCACGGATTGGCTGTTTACACAAAATGAAACGCTGGAATTCGACAGCCCGCGCTTT
CCGACCGCCCACACGCACGGCACGGGCTGCACGTTTTCCGCCTGCATCACCGCCGAGTTG
GCAAAAGGCTCGGACGTTTGCGAAGCCGTACAGACTGCCAAGGCCTACATCACGGCGGCA
ATCTCAAACCCTTTGGAAATCGGCGCAGGACACGGCCCGGTCAATCATTGGGCGTATCGG
GACTAACCGTAAAAATGCCGTCTGAAACAAAATGTTCAGACGGCATTTTTGAGGATTATT
CAGGCTTTTTCGCCAGCATCGTTACAAATTTAAACCGTATCGGATTGCCGTTTTCGTCTT
TGGCATGCATAGAACCCAATTCTTCTTTATATTCGACCAGTTCCCAATCCCGATAATAAT
CCTTCAGCTCGCCCTCTTTAAATTTAAAAGGGAACGGCATCGGACAGGGGAAATCCGCCG
TATCCATTGCCGATACAATCAAGTTGTACCCGCCCGCCGCCGTATGCGCCTGCATATCGG
CAATCACGTCGGGTACGCGCTGCGGCATCAGGAACATCAGCACCACTGTTGCCACAATAT
AATCAAACTCGCCCTGCAAGGCGGCGGCGTTCAAATCATATTCCAGCGTGCGGACGTTCA
AACCCTCCGCCTCTGCCAGCTCCGCCACGTTTGCCAAGGCGGCGGGATTGTGATCGACTG
CAGTAACTTCAAACCCCTTCAAACCGAGAAACAGCGCGTTGCGCCCCTGTCCGCAGCCCA
TATCCAACGCCCTGCCCGCCGGTACGGTATCCCGTGCCGCCGCGACCGCAGAATGCGTGG
CACTCATCCCGTATTTTTTGTGAAAATAGTCTGCCGCCGCGCAATACAGCGACAAACGGA
TTTCGGCATCGTCCGTTTTCGGTTTGACAGAAAACACCTGCTGCGGCGCAAACACACAAT
CGCCGCCGTCTGCCGACCAAACTTCTGCCGACCCGTCCGGTGCACGAACTTCGACATCGC
CCTGCAACACATTCAGGCAGACCCACTCCCCTTCCTCAGACGAATAGCCCGACAACAAAA

CTTCCGGCAGGTTTTCCACTTTCCATACAGGCATCTGTCCGAAACAAAACAACTCGCCAC
TTTGACCCACTATCCGCTCCTTCAwATTCAAAAATAAAGTTGCACATTATATGCCTATTT
TAATCCGCCGCAATCTTTCAGACGGCACGGCGCGCAAACCGCTTATAATCACGCCGGACA
CCACACAAAGGCACAATAATGAACCAAACCGTTTACCTTTACACCGACGGCGCGTGCAAA
GGCAATCCCGGCGCGGGCGGCTGGGGCGTGTTAATGCGCTACGGTAGCCACGAAAAAGAA
CTTTTCGGCGGCGAAGCGCAAACCACCAACAACCGCATGGAACTGACTGCCGTCATCGAA
GGACTGAAATCGCTCAAACGCCGCTGCACCGTCATCATCTGCACCGACTCGCAATACGTC
AAAAATGGCATGGAAAACTGGATACACGGTTGGAAGCGCAACGGCTGGAAAACCGCCTCC
AAACAGCCCGTCAAAAACGACGACTTGTGGAAAGAACTCGACGCTCTAGTCGGACGGCAT
CAAGTCAGTTGGACTTGGGTGAAAGGACACGCGGGACACGCCGAAAACGAACGCGCCGAC
GATTTGGCAAACCGTGGCGCAGCGCAGTTTTCCTGACTGCCGCTCCGGCAAAAATGCCGT
CTGAAACCGCTAATGGGCTTCAGACGGCATCGTCCTCCACCGTCATTCCCGCGCAAGCGG
GAATCCAAACCGTCGGGCAACGGCAATATTCAAAGATTATCTGAAAGTTTGAAGTTCTAG
ATTCCCGTTTTCACGGGAATGACGAAAAGTTGCAAGAATGACGGAGTTTCAGGCGGCATC
CGACCGCCCCGTCATTCCCGCGAAAGCGGGAATCTAAAAACCCAACGCTGCAAGATTTAT
CAGAAACAACTGAAACCGAACGGACTGGATTCCCGCCTGCGCGGGAATGACGGGATTTTA
GTAACCGTAGCAACCGCCTGCGCGACGGCTAAGGGGCTTCAGCAACCGTAGCAACTGCCT
GTGTGGGAATGACGGACAATGGGCTTCAGACGGCATCTCTTGCCTGCCGCTAAAACAGTT
TGCCGCACAACTGTTCAAACGCGTCCGATATGTTTCAACACACAGGACGACACATAAAGC
ACCTCCCTATGTGTCGTCCTGATTTGGAAGGGGTTACACCCCCTCCCAAATAAAGTCTGA
TCCTGCCGCCCTAAAGGGCGGGGTTTCAACCGAAAAGGAAATACGATGAAGTGGTACAAT
TAGCGGCAATGCGGACAGACAAATTAAACTATAGTGGATTAAATTTAAACCAGTACGGCG
TTGCCTCGCCTTAGCTCAAAGAGAACGATTCTCTAAGGTGCTGAAGCACCAAGTGAATCG
GTTCTGTACTATTTGTACTGTCTGCGGCTTCGTTGCCTTGTCCTGATTTTTGTTAATCCG
CTATATCAGAAATTACCCTACCGTTTTTTAAACACTTTCAGGAATAAGGAAAAATGACCG
CCCAACCCTGCCCCATCTGCACGGCGCAAAATGAAGACGTTTTGCTGCAAACCCCCAACC
TCCGCGTCATCGCCGTCCATAACGACAGCGGTTCGCCTGCATTCTGCCGCGTCATTTGGC
GTAAGCATATTGCCGAAATGACCGACCTTTCGGCAGCGGAACGCGGCGAATTGATGGAAA
TGGTGTACAAAGTCGAAGCCGCTATGCGCCAAGTGTTCCGGCCGGCAAAAATCAACCTCG
CCAGCTTGGGCAATGTCGTGCCGCACCTGCATTGGCATATTATCGCCCGCTTTGAAAACG
ATGCGTCTTTCCCCGCGCCGATTTGGGCAAACCCCGTCCGGAAACACGGTATGACCCTGC
CGCAAGATTGGACGGAACAGCTTAAAAAGCTGCTTTAAGCCCGCCGATGCCGTCTGAAAC
CGTATGAAAGGGAAATTATGACCGAACCGACCTCCCGCCGCCGTTTTCTGAAAACCTGCA
CCGCCGCTGCCGGCGCGGGGCTGCTTCAGGCTTGCGGCACATCCGCCACATCCGTTCCGC
CCCTTCCCTCTTCCCATTCCGTTGTGAAAGCCCGAACCGTGCCTCTCCAAACGCCACGCC
GTCAAAGTTCGGACGGCAACCTTCTGCGCGTTGTCGCTTCGTCAGGATTTGCCGAAGACA
CCAACCGCGTCAACACAGCCTTAACCCGCCTTTACAATGTCGGTTTTACCGTAACCAACC
AACAGGCGGGCAGCCGCCGTTTCCAACGGTTTGCCGGCACGGACACGCAACGTGCCGCCG
ATTTCCAAGAGGTCGCTTCCGGCCGCGTCGCCACGCCTAAAGTGCTGATGGGTTTGCGCG
GCGGTTACGGTGCGGCGCGGATTCTGCCGCATATCGATTTTGCTTCGCTCGGCGCAAGGA
TGCGCGAACACGGCACGCTCTTTTTCGGATTCAGCGACGTATGCGCCGTCCAGCTGGCAT
TGTTGGCAAAAGGCAATATGATGAGTTTTGCCGGCCCGATGGCTTATAGCGATTTTGGCA
AACCCGCCCCCGGTGCGTTTACGATGGATGCCTTTATCAAGGGTGCAACCCAAAACCGCC
TGACCGTTGATGTTCCTTATATCCAACGCGCCGATGTCGAAACCGAAGGCATATTGTGGG
GCGGCAACTTAAGCGTCCTCGCCTCGCTCGCCGGCACGCCTTATATGCCCGACATCGACG
GCGGCATTTTGTTCCTCGAAGATGTCGGCGAACAGCCCTACCGCATCGAACGTATGCTCA
ATACGCTGTATCTTTCGGGTATTTTGAAGAAACAGCGCGCCATCGTGTTCGGCAATTTCC
GTATGGAAAAAATTCGAGATGTCTATGATCCGTCTTATGATTTTTCTGCCGTTGCCAACC
ATGTTTCGCGCACGGCGAAAATCCCCGTGCTGACGGGCTTCCCGTTCGGACACATTGCCG
ACAAAATCACTTTCCCTCTAGGCGCGCACGCCCGAATCCGTATGAACGGAAACAGCGGTT
ATTCGGTCGCGTTTGAAGGCTACCCCACACTCGATGCGTCCGCCCTGACTTTGGATACCC
TGCTCCCACCGCCGGATTTGCCCATCTTCCCCGAAAGCGGTGTTGCCGATATTTCGGAAT
AAACCCGCAAACGGACAAATGCCGTCTGAAGCCTTCAGACGGCATTTCCCAAGACGGCGG
CAGATTACAGCAATGCCCGAATATCGGCTTCGATTTCTTCGGGCGTAACACTAGGCGCAA
AACGCTCGACCACTTCGCCGTCGCGGTTGACGAGGAATTTGGTAAAGTTCCATTTGATGT
CGCCTTCGTCGCGCTTCTCTCCCAAAGCTGCGAGCTTCAACACGAAATCTTTAAACAGAT
GATTGCCTTTATCTTGCGGTTTGACGGATTTCAGGTAGGCATACAAGGGCGCGGTATTTG

CTCCATTGACTTCGATTTTGTCGAAAATCTTAAACTTCGTGCCAAACTTCATCATACACA
CTTGGGCAATTTCTCCGCTGCTTTCGGGAGCCTGTTCGCGGAACTGGTTGCACGGAAAAT
CCAAAATCTCCAAGCCTTCTGCGGTATATTGTGCATACAGCTTCTGCAAAGCCTCGTATT
GCGGGGTCAGACCGCAACGCGTTGCCGTGTTGACAATCAGCAGAACCTTGCCGCGATAGC
CTGACAAATCAACCGCATTGCCTTCTGCATCTTTCATTTGAAAATCGTAAATACCCATTT
TTATCCTTATCTGATGTAAACCGATGCCATCTGAAACGTGCTTCAGACGGCATGAAAGCA
GCAATTGTATAGCCGATTAAAATAAAAAATCCACATCCTTTTCCATTCCCGTCCCAATCC
GCAATAAAAAACTGCACCCGAAAACGGGTGCAGTTGCTCATTTCATACCGCAAAACTTAT
TTGTCGCGGCCGAATACGATTTTAGTGGCTTGGATGGCGACACAGATTGCACCGCCGATA
AAGACCAAGTCAGCTGCCGTACGTACCCAACGCAAGGTATCGAGGATTTCCATTTGCAGG
AACTCTTCGCTGCGGGCATACCACAGACCGTGCGTGATGGAGGCGTATGCCTGAATCGCG
CCGACAGGCAGCAGGCTGATGGCAATCATACCGGCCAAGCCGCCGTTGAGCAGCCAGAAG
CCCCAAGTCATCAGTTTGTCGTCAAACTGCGCGTTCGGTTTCAAATAACGGGCAACCAGC
AATACGAAGCCCAATGCCAAGAAACCGTACACACCAAACAAGGCGGCGTGCGCGTGAACG
GCAGAAGTGTTCAAACCTTGGATATAGAACAGGGAAATCGGCGGATTGATCAGGAAGCCG
AATACGCCGGCACCGATCATATTCCAAAAGGCGACTGCCACGAAGCACATCAGCGGCCAA
CGCAGGCGTTTCGCCCAGTCGGACAGGTGTTGGTAAGACCAGTGTTCGTATGCTTCACGG
CCCAGCAACACCAGCGGCACGACTTCCAAAGCGGAGAAGCAGGCACCGATTGCCATAGAG
GCGGAGGTAGAGCCGGAGAAGTACAGGTGGTGCAGCGTGCCCGGAACGCCGCCCAACATA
AAGATGGCGGCAGCGGCCAAAGTGGAGGCAGTGGCGGTACTGCGGCGGACAAAGCCCATA
TTGTAGAAGACAAAGGCAAAGGCGGCAGTGGCAAATACTTCGAAGAAGCCTTCTACCCAC
AGGTGAACCACCCACCAACGCCAGTATTCCATAACGGCAATCGGGGATTTTTCGCCATAG
AACAGGCCTGGTGCGTAGAATACGCCCACACCGACCATAGAAGCTACGAAGATAGCCAAC
AGGTTTTTGTCCACGCCTTTTTCTTTAAAGGCGGAAACCGTGCAACGCAACATCAGGAAC
AGCCATAACAGCAGACCGACCATCAAAAGGAGTTGCCAGAAACGTCCCAAATCGAGGTAT
TCGTAACCTTGGTGTCCGAACCAGAAGTTAAATTCCGGGGGAAGGATGTGCGTCAACGCG
AAGAAGTTGCCCGCGTAAGAACCGCCGACCACGATGAAGAGGGCGATATAGAGGAAGTTT
ACGCCGGCACGTTGGAACTTGGGATCTTTACCGCCGTTGACAATCGGCGCGAGGAACAAA
CCTGCCGTCAAAAAGCCGGTTGCAATCCAGAAGATGGCGGATTGGATGTGCCAAGTACGG
GTCAGGGCGTAGGGGAACCAGTCGGACATTTCAAAGCCCAACGCCTCGTCAATGCCGTAG
AAACCCTGGCCTTCGACGGTGTAGTGCGCGGTCAGTCCGCCCAGCAATACTTGTACCACA
AACAGGGCGACCGTCAGGAAGACGTATTTGCCCAATGCTTTTTGCGAAGGGGTCAGTTGG
ATTTTGGAAATCGGGTCTTCAGACGGCACTTCCACTTCCTCGTGTTTGGTCAGGAAGGAA
TAACCCCACATCAGCAAACCGATGCCCATCAGCAGAAGAACAACGCTGGTGAATGACCAC
ATATAGTTTTCAGTGGTCGGTACGTTGTTGATCAAAGGTTCGTGCGGCCAGTTGTTGGTG
TAAGTAAAATTCTCGTCAGGACGGTTGGTCGAAGCAGACCAAGAAGTCCAGAAGAAGAAG
TTGAACAGTTTTTCACGCGCTTCTTGGCTTGGCAATGTATTGTTTTTCATTGCAAAGTGT
TCGCGAGTGGTTTGGAACTTAGGATCGTCGCTGTACACACCGTGGTAGTAAGGCAGGATG
CTTTCGATGGCTTTCACGCGCGTATCGCTGATGACGACGCTGCCGTCTTCCTTCACGCGG
CTTTGATTGCGGTATTCGTCGGCCAGGCGTGTTTTCAAGACGGCTTGTTCCTCGGGGGAA
ACCTCGTCGAATTTTTTGCCGTAAGTCTGTTGCGCGGTCAAATCCAACCAGGCAACCAAC
TCACGATGCAGCCAGTCCGCCGTCCAGTCCGGAGCCTGATATGCACCGTGACCCAAAATC
GAACCGACTTCCATACCGCCGGTAGTCTGCCATGCAGACTGACCTGCCAAAATATCGTCT
TTCGTCATCAAGACCTTGCCGGATGCGGAAACGACCTGTTCGGGGTAAGGCGGGGCTTTT
TTGTAAACCTCGCTGCCCATATAGCCAAGAATGGTAAAGCATAcCGGGTAC

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 2>:

gnm_2

CGAGGCGCAGATACAGGTTTTGGAAGATGTGCACGTCAAGGCGAAGCGCGTACCGAAAGA
CAAAAAAGTGTTTACCGATGCGCGTGCCGTATCGACCCGTCAGGATATATTCAAATCCAG
CGAAAACCTCGACAACATCGTACGCAGCATCCCCGGTGCGTTTACACAGCAAGATAAAAG

CTCGGGCATTGTGTCTTTGAATATTCGCGGCGACAGCGGGTTCGGGCGGGTCAATACGAT
GGTGGACGGCATCACGCAGACCTTTTATTCGACTTCTACCGATGCGGGCAGGGCAGGCGG
TTCATCTCAATTCGGTGCATCTGTCGACAGCAATTTTATTGCCGGACTGGATGTCGTCAA
AGGCAGCTTCAGCGGCTCGGCAGGCATCAACAGCCTTGCCGGTTCGGCGAATCTGCGGAC
TTTAGGCGTGGATGACGTCGTTCAGGGCAATAATACCTACGGCCTGCTGCTAAAAGGTCT
GACCGGCACCAATTCAACCAAAGGTAATGCGATGGCGGCGATAGGTGCGCGCAAATGGCT
GGAAAGCGGAGCATCTGTCGGTGTGCTTTACGGGCACAGCAgGCGCAGCGTGGCGCAAAA
TTACCGCGTGGGCGGCGGCGGGCAGCACATCGGAAATTTTGGCGCGGAATATTTGGAACG
GCGCAAGCAGCGATATTTTGTACAAGAGGGTGCTTTGAAATTCAATTCCGACAGCGGAAA
ATGGGAGCGGGATTTACAAAGGCAACAGTGGAAATACAAGCCGTATAAAAATTACAACAA
CCAAGAACTACAAAAATACATCGAAGAGCATGACAAAAGCTGGCGGGAAAACCTGGCAAC
CGCAATACGACATTACCCCCATCGATCCGTCCAGCCTGAAGCAGCAGTCGGCAGGCAATC
TGTTTAAATTGGAATACGACGGCGTATTCAATAAATACACGGCGCAATTTCGCGATTTAA
ACACCAAAATCGGCAGCCGCAAAATCATCAACCGCAATTATCAGTTCAATTACGGTTTGT
CTTTGAACCCGTATACCAACCTCAATCTGACCGCAGCCTACAATTCGGGCAGGCAGAAAT
ATCCGAAAGGGTCGAAGTTTACAGGCTGGGGGCTTTTAAAGGATTTTGAAACCTACAACA
ACGCGAAAATCCTCGACCTCAACAACACCGCCACCTTCCGGCTGCCCCGCGAAACCGAGT
TGCAAACCACTTTGGGCTTCAATTATTTCCACAACGAATACGGCAAAAACCGCTTTCCTG
AAGAATTGGGGCTGTTTTTCGACGGTCCTGATCAGGACAACGGGCTTTATTCCTATTTGG
GGCGGTTTAAGGGCGATAAAGGGCTGCTGCCCCAAAAATCAACCATTGTCCAACCGGCCG
GCAGCCAATATTTCAACACGTTCTACTTCGATGCCGCGCTCAAAAAAGACATTTACCGCT
TAAACTACAGCACCAATACCGTCGGCTACCGTTTCGGCGGCGAATATACGGGCTATTACG
GCTCGGATGACGAATTTAAGCGGGCATTCGGAGAAAACTCGCCGACATACAAGAAACATT
GCAACCGGAGCTGCGGGATTTATGAACCCGTATTGAAAAAATACGGCAAAAAGCGCGCCA
ACAACCATTCGGTCAGCATTAGTGCGGACTTCGGCGATTATTTCATGCCGTTCGCCAGCT
ATTCGCGCACACACCGTATGCCCAACATCCAAGAAATGTATTTTTCCCAAATCGGCGACT
CCGGCGTTCACACCGCCTTAAAACCAGAGCGCGCAAACACTTGGCAATTTGGCTTCAATA
CCTATAAAAAAGGATTGTTAAAACAAGATGATACATTAGGATTAAAACTGGTCGGCTACC
GCAGCCGCATCGACAACTACATCCACAACGTTTACGGGAAATGGTGGGATTTGAACGGGG
ATATTCCGAGCTGGGTCAGCAGCACCGGGCTTGCCTACACCATCCAACATCGCAATTTCA
AAGACAAAGTGCACAAACACGGTTTTGAGTTGGAGCTGAATTACGATTATGGGCGTTTTT
TCACCAACCTTTCTTACGCCTATCAAAAAAGCACGCAACCGACCAACTTCAGCGATGCGA
GCGAATCGCCCAACAATGCGTCCAAAGAAGACCAACTCAAACAAGGTTATGGGTTGAGCA
GGGTTTCCGCCCTGCCGCGAGATTACGGACGTTTGGAAGTCGGTACGCGCTGGTTGGGCA
ACAAACTGACTTTGGGCGGCGCGATGCGCTATTTCGGCAAGAGCATCCGCGCGACGGCTG
AAGAACGCTATATCGACGGCACCAACGGGGGAAATACCAGCAATTTCCGGCAACTGGGCA
AGCGTTCCATCAAACAAACCGAAACTCTTGCCCGCCAGCCTTTGATTTTTGATTTTTACG
CCGCTTACGAGCCGAAGAAAAACCTTATTTCCGCGCCGAAGTCAAAAATCTGTTCGACA
GGCGTTATATCGATCCGCTCGATGCGGGCAATGATGCGGCAACGCAGCGTTATTACAGCT
CGTTCGACCCGAAAGACAAGGACGAAGACGTAACGTGTAATGCTGATAAAACGTTGTGCA
ACGGCAAATACGGCGGCACAAGCAAAAGCGTATTGACCAATTTTGCACGCGGACGCACCT
TTTTGATGACGATGAGCTACAAGTTTTAAAGGCAGCCCGCATTTTGTAGAAAACCGCAAT
GCCGTCTGAAAGCCCTTCAGACGGCATTTGTTTCCCCAAACGCATCATCCTGCCGCAAGC
CTATGCCAATCCGTTTTATCGCATCGGCAACTCAAAGAAAAATCCATTTCATTCCCACGC
AGGGAAGCCGGTTTTTGATTTCGGTTATTTTTGGTTGTTTCGGGTAATTTATGAGTCGTC
ATTCCCGCAAAAGCGGGAATCAGTTTTTTTAAGTTTCAGCCATTTCCGATAAATTCCTGT
GGCTTTAGCTTTCCGGATTCCCACTTTCGTGAGAATGACGTGGTGCAGGTTTCCGTACGG
ATGGATTCGTCATTCCCGCGCAGGCGGGAATCTAGACCGTTCGGTTTCGGTTTTTTTGGT
TAGTGCCGCAACATTAAATTTCTAGATTCCCACTTTCGTGGGAATGACGGCGGAGCGGTT
TCTGCTTTTTCCAATAAATGCCCCCAACCTAAAATCCGTCATTCCCGCGCAGGCGGGAAT
CTAGACATTCAATGCTAAGGCAATTTATCGGAAATGACTGAAACTCAAAAAACTAGATTC
CCACTTTCGTGGGAATGACGTGGTGCAGGTTTCCGTATGGATGGATTCGTCATTCCCGCG
CAGGCGGGAATCTAGTCCGTTCGGTTTCGGTTTTTTTGGCTAATGCCGCAACATTAAATT
TCTAGATTCCCACTTTCGTGGGAATGACGGCGGAGCGGTTGCTGTTTTTCCCAATAAATG
CCCCCCAACCTAAAATCCGTCATTCCCGCGCAGGCGGGAATCTAGTCCGTTCGGTTTCGG
TTTTTTTGGCTAGTGCCGCAACATTAAATTTCTAGATTCCCACTTTCGTGGGAATGACGG
CGGAGCGGTTTCTGCTTTTCCCAATAAATGCCCCCAACCTAAAATCCGTCATTCCCGCGC

AGGCGGGAATTTAGACATTCAACGCTAAGGCAATTTATCGGAAATGACTGAAACTCAAAA
AACTGGATTCCCTCTTTCGTGGGAATGACGTAGTGCAGGTTTCCGTACGGATGGATTCGT
CATTCCCGCGCAGGCGGGAATCTAGACATTCAATGCTAAGGCAATTTATCGGAAATGACT
GAAACTCAAAAAACTGGATTCCCGCTTTCGTGGGAATGACGCGATTAGAGTTTCAAAATT
TATTCTAAATAGCTGAAACTCAACGCACTGGATTCCCGCCTGAGCGGGAATGACGAAGTG
GAAGTTACCCGAAACTTAAAACAAGCGAAACCGAACGAACTGGATTCCCACTTTCGTGGG
AATGACGGAATGTAGGTTCGTGGGAATGACGGGATGCAGGTTTCCGATGGATGGATTCGT
CATTCCCGCGCAGGCGGGAATCTAGACATTCAACGCTAAGGCAATTTATCGGAAATGACT
GAAACTCAAAAAACTGGATTCCCACTTTTGTGGGAATGACGCGATTAGAGTTTCAAAATT
TATTCTAAATAGCTGAAACTCAACGCACTGGATTCCCGCCTGAGCGGGAATGACGAATTT
CAGGTTGCTGTTTTTGGTTTTCTGTTTTTGTGAAAATAATGGGATTTTAGCTTGTGGGTA
TTTACCGGAAAAAACAGAAACCGCTCCGCCGTCATTCCCGCGCAGGCGGGAATCTAGTCC
GTTCGGTTTCGGTTTTTTTGGCTAGTGCCGCAACATTAAATTTCTAGATTCCCACTTTCG
TGGGAATGACGGGATGTATAGTGGATTAACAAAAACCAGTACGGCGTTGCCTCGCCTTAG
CTCAAAGAGAACGATTGTCTAAGGTGCTGAAGCACCAAGTGAATCGGTTCCGTACTATTT
GTACTGTCTGCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATAAATTTAATC
CACTATATTTTTTGTTCCAAAGTCAAAATATGCCGTCCGAACATTCGGGCGGCAGACAAA
ACGGCACTGCCCGATAAAGGCAGTGCCGTTGTCCGTTTCAAACCGTGAAACATCAGCCCA
AATTAAAGGCTTTATGCAATACCCTGGTTGCCAGTTCCATGTATTTTTCATCAATCAATA
CGGAAACTTTGATTTCGGAGGTGGAAATCATTTGGATGTTGATACCCTCTTCGGCGAGCG
TGCGGAAGATTTTGGCGGCTACACCGACGTGCGAACGCATACCCAAACCGACTGCGGAGA
CTTTGCATACGGTGTCGTCGCCATCAATAGAAGCCGCGCCGATACTGTCTTGGCGTTCCG
ACAGGATTTCCAAAGTCTGCTTGTAATCGCCGCGCGGTACGGTAAAGGAAAAATCGGTTG
TGCCTTCGCTGCCGACATTTTGGATAATCATATCGACTTCGATGTTGGCATCGGCAACCG
CGCCTAAAATCTGATAGGCGACGCCAGGTTTGTCGGGTACGCCGCGCACGTTGATGCGGG
CTTGGTTTTTATCGAATGCGATACCGGTTACGGCAGCTCTTTCCATGTTGTCGTCCTCTT
CAAAGGTAATTAAGGTGCCATTGCCGCCGTCTTGCAGGCTGCTCAGTACGCGCAGGCGCA
CTTTGTATTTTCCGGCGAATTCTACTGAACGGATTTGCAAAACTTTCGAACCGAGGCTTG
CCAGTTCGATCATTTCTTCAAATGTAACCGTATCCATGCGGCGCGCTTCGGGTACGACGC
GGGGGTCGGTTGTGTAAACGCCGTCTACGTCGGTATAGATTTGGCACTCGTCGGCTTTGA
GCGCGGCGGCAAGCGCGACGGCGGAAGTGTCGGAACCGCCGCGTCCGAGCGTGGAAATAT
CGCCTTCACTGCTGATGCCTTGGAAGCCGGCAACGATGACGACTTTGCCGGCGGTAAGGT
CGGCACGCATTTTTTCGTCATCAATGCTTTCGATGCGGGCTTTGGTGTGGGCGGTATCGG
TTTTGAGGGCGACCTGCCAGCCTGTGTAGCTTTTGGCATCCACGCCGATGTCTTTCAATG
CCATCGCCAAAAGGCCGATGGTTACTTGTTCGCCGGTAGCTAAGACGACGTCCAGCTCGC
GCGGATCGGGATGCTCTTGCATTTCGTGCGCCAGTGCGACCAGTCGGTTGGTTTCGCCGC
TCATGGCGGATACGACGACTACGATGTCGTGTCCTTCGGCGCGGGCTTTGGCGACACGTT
TGGCTACGTTTTTGATGCGTTCGGGCGAGCCTACTGATGTGCCGCCGTATTTATGTACGA
TTAACGCCATGTTTCGTGCTTTCTTGTGGGGGTTGTCGGGCAGCTTGGTTTGCTGGAAAA
AGGGTTATTATTACTATTTTTTACATGGAATTCAAGAACGGACTGCGCTTTCCCGCCTGC
CGTTTGACAGCGGTCAGCGAAAAACCTGTTCTTTCAGATTGTTGACAAAATGCCGTCTGA
ACGGTTTTCAGACGGCATCCGGACGACAATCAGGCGGCGGACAACGCATTTTGCTGGTGT
TGCAGCAGTTCGCCTATGCCTTTTTGCGCCAGTGCAACCAGTTTGCCCAATTCGTCCAAA
CTGAACGGCGCGTCTTCCGCCGTCCCCTGTATTTCGATGATTTTTCCCGATGCGGTCATG
ACGATATTCACATCACTGTCGCAACCGGAGTCTTCGGGATAATCCAAATCCAAAAGCGGC
ACGCCGTTCACTACGCCTACTGACACAGCGGCAACGGCTTCGCGGATGGGGTTTTCACTC
AAAATGCCGTCTGAAACCAGTTTGCCGACGGCGATTTGCAGCGCGACAAACGCACCGGTA
ATCGAAGCCGTGCGCGTACCGCCGTCTGCCTGAATCACATCGCAGTCAATCAAGATTTGT
CGTTCACCGAGTTTTTCCATATCCACGACCGCGCGCAGGGAACGCCCGATCAAACGTTGG
ATTTCTTGTGTGCGCCCGGACTGTTTGCCCGCCGAAGCTTCGCGGAGCATCCGGGAAGCA
GTTGAGGCAGGCAGCATCCCGTATTCCGCCGTTACCCAGCCTTGGTTTTTACCGCGCAGA
AACGGCGGGACGTTTTCATCTATGGAAGCGGTACAAATCACTTTGGTATTGCCGCATTCA
ATAAGGCACGAACCGTCCGTATGCGGCAGGAAATGAGGGGTGATTTTGATATCGCGCAGG
CTGTCGGCGGCGCGCGAGATGCGGATGTAATCAGGCATACTGCCCTCCCGTTAAAAACAG
ATAAATTAAAAAGCCTTAAATATGAAAAATCACATTTAAGGCCTTCAAACTGAAAATTTC
TACGCCTCTTCGGCTTTGCTGCGGATAATCAAAAGCGGCAGGTGGCTTTGGCGCATTACC
GTTTCGGCAAAACTGCCCATTAAAAGGTGCATCAGCCCGGTACGTCCGTGCGTACCCAAC

ACCAGCAGGTCGGCACCGTTTTCATCGGCATAATCAACCAAATCCTGCGCCATTTCACGC
GCACCCTTATTGGCAACCAGCAGGTGTTTGACGGTATTTTCCACACCCAGTTCCTGGGCG
GTGCGCTCGGCGGCATCCAAAACTTCGTTGCCTTGCGCGACGGCGGCGGCTTCGTAGCTT
TCGTGTTGCAAAAATTCGGGGGCGAGTGCCATATATTCGGCAGGATTGGCAACGTGCACC
AAAGTCAGGCGCGCACCGTTGACCCCGGCAAGCTCGGCGGCATGTTTCAGGGCATTGATG
GACGTTTCACTGCCGTCAACGGCAACAACCAAATGTTTGTACATATCGTATTCTCCTTTT
GCACCGCCTCGCGGTGCCCTCTTGTCGGATGGGCGCAGGGACAGTTTGCGCTGTTTCATT
ATAGACCCGCCGTCGGGCTTTATACAACAGCCGAACAGCCCGACCGCTTTCCAGTATAAT
ATGCCGCTTCCGTGCAGTCAGGCATTTTTTGCCGGCTTTCGTTCACTTTTTGATTTGACG
CAATCTTGCAGGATTCGACCATGTCCGACAACGCTTTGACCTCTTCGCGACGCTTCGGCG
GCATCGCCAGACTCTACGGAGACTCTGCCTTGGCGCACTTTTCACAGGCACACGTCTGCG
TAGTCGGCGTGGGCGGTGTCGGCTCGTGGGCGGTCGAGGCTTTGGCGCGGACGGGCATCG
GACGTTTGACTTTGATTGATTTGGACAACGTTGCCGAATCGAATGTCAACCGCCAGCTGC
ACGCCCTGACCGGCGACTTCGGCAAAGCAAAAGTTACCGCCTTGCGCGAACGCATTACAC
AAATTAATCCGCAATGCGAAGTGTTTGAAATTGAAGATTTCGTTACCGAAGACAATTTGC
CGGAATACTTCGGAAAAGGTTTTGATTTCGTCATCGACGCGATCGACCAAGTGCGCGTCA
AAGCAGCAATGGCGGCTTATTTTGTGGAACGCAAACAACCGTTTGTCCTCAGCGGCGGCG
CGGGCGGACAAAAAAATCCGGCGTTAATCCAAACCGCCGATTTGAGCCGCGTAACCCACG
ACCCGCTGCTTGCCAACCTGCGCTACACCTTGCGGAAACGCTACGGATTCAGCCGCGATA
CGAAAGCAAATATGCGCGTGCCTTGCGTGTATTCGACCGAAAATATCGTGCCGCCGCAGT
CTAGGGAGGCTTGTTCGGCAGATGCCGCTCCGCAAGGCTTGTCGTGCGCCGGCTACGGTG
CAAGCATGCTCGTTACCGCTTCGTTCGGGCTATATTGCGCACAGGCGGCGGTGGAACACA
TCGCAGACAAAAAATAAGCAATGCCGTCTGAAACAGGATTCAGACGGCATTTGAACAAAC
TATGGTTATGATTTAAGACAACAAAGGATACGGATAAAAAATAACATAAAATATATGATT
CCTAATAATATACCAAGTATCGGAGAGCTATTTAATGGAATTCGTTAATAATTTAGTTAT
TTTTTCATTTTTATTACTAATGCTTATTCCGATATTTTTTGTAGTATATGGTATATACCA
TAAGATACGTTATCGCAAAATATGTATCCTAAGAACAAGTTTTATATTATTAGTGGTAAT
ACTTTGCAGTATGTATTACATATATTGCCGTTATCTTGACCAACAAAAAGTAGCTTATTA
TTGCATAGATGAACAATGTATTTCTATTGTTCATCTATACAAAGATTATGGTATAAACTC
TCCCACATATGCGAGAATTTACGCAGGAAAAATATTGTTTAGATTTCAAGTAAGAGCTAA
AAATTACGCTGAATTACTTATGGAAGATGATATATCAATTAGTAAAAAAATTTTGGGGAA
TAAATTTATCATTTATGGGTCGCTACCTGTAATATACGGTAATGTAGATAATATTGAAGT
AAAAGAAGCTACTGGTTATATAGATAGATCCAGTACTGATTATATTGTCTCAAGAAACTT
AAAATTCAGACATTTATATTAATTAAGAGGTTTTAGCAAGAGTGCCGTCAAAATATAGGG
CGCATCATCGAATTCGCGAAAGACAAACGCTACGATGAACGTTTCAAGGATTTGAAAAAA
GAATCCATAGGCTATCTGAACCGGCATCCCGGTTTGGTGTCCGACTACCTGAAGGCGGCA
ATCAAGCTGTCGGTTCAGAAAAACCAACATCAGCACGCCTAAAACCGTATTCACAACCTG
CTCCTTTTCAAAACATTTGCATTTAAAAGCCGTTATAATGCCGTCTGAACATCTGCCCGA
CCACATTATACGTGAATGTCGGCAGATTGTTTTCTTTTGTAAACTTATATTAAAATCCAC
TTACCGATTCACGCCATGCCGCCCATCCCTGCCCCATCTGCACCATCCGAGCACACTGTC
GCATGGGTATTCGGCCAACCCGTTACCGATTTGCCCCAGGATTTGTTTATTCCGCCCGAT
GCATTGAAAGTCGTATTGGGCAGCTTCCAAGGCCCTTTGGATCTACTGCTGTATCTGATC
CGCAAACAGAATATCGACGTACTGGATATTCCGATGGTGAAGATTACCGAGCAGTATCTG
CACTACATCGCCCAAATAGAAACCTATCAGTTTGATTTGGCGGCGGAATATCTTTTGATG
GCAGCAATGCTGATTGAAATCAAATCGCGCCTGCTGCTGCCGCGTACCGAAACCGTCGAA
GACGAAGAAGCCGACCCGCGTGCCGAGTTGGTGCGCCGCCTGCTGGCTTACGAACAGATG
AAGCTGGCGGCGCAGGGTTTGGACGCGCTGCCCCGAGCCGGACGGGATTTCGCGTGGGCT
TACCTGCCGCTGGAAATTGCCGTCGAAGCCAAGCTGCCCGAAGTCTATATTACCGACTTG
ACGCAAGCGTGGCTGGGTATTTTGTCTCGGGCAAAACACACGCGCAGCCACGAAGTAATC
AAAGAAACCATCTCCGTGCGCGCGCAAATGACGGCAATCCTGCGCCGTTTGAACGGACAC
GGAATATGCAGGTTTCACGACCTGTTCAATCCCAAACAGGGCGCGGCTTACGTGGTCGTC
AACTTCATCGCACTGTTGGAGCTTGCCAAAGAAGGATTGGTCAGAATCGTGCAGGAAGAC
GGTTTCGGAGAAATCCGAATCAGCCTCAATCATGAGGGGGCGCATTCAGACGGCATTTCC
GGCACACGAGGCGGGCGCGATGTGTTCTAATACGCCCCAAGCCGCCACCAAAAATCGGGA
GACACGCCATATGACCGGCATCATACATTCGCTGCTTGACACCGACCTCTACAAATTCAC
TATGCTGCAAGTGGTTCTGCACCAGTTTCCGCAGACGCACAGCCTTTACGAATTCCGCTG
CCGCAACGCCTCGACCGTCTATCCGCTTGCCGACATCAGGGAAGACTTGGAAGCCGAACT

CGACGCGCTCTGCCAACTACGCTTCACCCACGACGAACTCGGCTATCTGCGCTCCCTGCG
TTTCATTAAAAGCGACTTTGTCGATTATCTCGAACTCTTCCAGCTCCAACGCCGCTTTGT
CGAAATCGGCACAGACGATAAAGACCGTCTGAACATCCGCATCGAAGGTCCGATGATACA
GGCGATGTTTTTTGAAATCTTCATCCTCGCCATTGTCAACGAACTTTACTTCCGCCGCCT
GGAAACCCCTGCAGTCATAGAAGAAGGCGAACGCCGGCTTCAAGCCAAAGCCGCGCGCCT
CAAAGAAATCGCCGCCGCACAAAACCCCGACGAACCGCCCTTCCTGATTTCCGACTTCGG
CACGCGCCGCCGCTACAAGCTCGCGTGGCAGGAACACGTCATCCGCACCCTGCTTGAAGC
CGCCCCCGGCATCGTACGCGGCACCAGCAATGTCTTTCTCGCCAAAAAACTCGGCATCAC
CCCCATCGGCACCATGGCGCACGAGTTCCTGCAGGCATTCCAGGCCCTCGACGTACGCCT
GCGGAATTTCCAAAAGGCCGCGCTCGAAAGCTGGGTGCACGAATACCGGGGCGATTTGGG
CGTTGCCCTGACCGACGTGGTCGGTATGGATGCCTTCCTGCGCGATTTCGACCTCTATTT
CGCCAAACTTTTCGACGGGCTGCGCCACGACAGCGGCGACCCTTACGTTTGGGGCGACAA
AGCCTACGCCCACTATCAAAAGCTCAAAATCGACAGCCGCACCAAAATGCTGACCTTCTC
CGACGGGCTGGACATCGAACGCTCTTGGGCATTGCACCAATATTTCAAAGACCGCTTCAA
AACCGGCTTCGGCATCGGCACCAACCTCACCAACGATATGGGGCATACGCCCTTGAATAT
CGTCTTGAAACTGGTCGAATGCAACGGGCAGTCCGTCGCCAAGCTGTCCGACTCTCCGGG
CAAAACCATGACCAACAACAGCACCTTCCTCGCCTACCTGCGCCAAGTGTTCGACGTACC
CGAACCCGAAACGCCGTAAACCGGCAGAAAAAGCGCACAATTCCTGTTTCTGCCGCATAA
AATCTTTTAAAATACCGCCTGATTTGAATTTAACCGAAAGACCGAACTTCATGAACCTAC
ATCAAACCGTCGAACACGAAGCCGCCGCCGCCTTTGCCGCCGCAGGCATCGCCGACAGCC
CTATTGTTTTGCAGCCGACCAAAAACGCCGAACACGGCGATTTCCAAATCAACGGCGTGA
TGGGTGCGGCGAAAAAAGCCAAACAAAACCCGCGCGAGTTGGCGCAAAAGGTCGCCGAAG
CATTGGCGGACAACGCCGTGATTGAAAGCGCGGAAGTCGCCGGTCCGGGCTTCATCAACC
TGCGCCTGCGCCCCGAATTTCTCGCGCAAAACATTCAGACGGCCTTGAACGACGCTCGTT
TCGGCGTGGCAAAAACCGACAAACCGCAAACCGTCGTTATCGACTATTCTTCGCCCAATC
TGGCGAAGGAAATGCACGTCGGCCACCTGCGTTCCAGCATCATCGGCGACAGCATTTCGC
GCGTGTTGGCATTTATGGGCAATACCGTTATCCGTCAAAACCACGTCGGCGACTGGGGTA
CGCAGTTCGGTATGTTGGTCGCTTATTTGGTCGAGCAGCAAAAAGACAATGCCGCGTTCG
AGCTGGCGGATTTGGAGCAGTTTTACCGCGCCGCCAAAGTGCGCTTTGACGAAGACCCTG
CCTTTGCCGACACCGCACGCGAATACGTTGTGAAGCTGCAAGGCGGCGATGAAACCGTTT
TGGCATTGTGGAAACAGTTTGTCGATATTTCGCTCTCGCACGCCCAAGCCGTTTACGACA
CGCTGGGCTTGAAGCTGCGTCCTGAAGACGTGGCAGGCGAATCGAAATACAACGACGATT
TGCAGCCCGTGGTCGATGATTTGGTTCAAAAAGGTCTGGCGGTTGAGGACGACGGCGCGA
AAGTCGTGTTCTTGGACGAATTTAAAAACAAAGAAGGCGAACCCGCCGCATTTATCGTGC
AAAAACAAGGCGGCGGCTTCCTCTACGCCTCCACCGATTTGGCGTGCCTGCGCTACCGCA
TAGGCCGTCTGAAAGCCGACCGCCTGCTGTACGTCGTCGACCACCGCCAAGCCCTGCACT
TCGAACAACTTTTCACCACTTCCCGCAAAGCAGGCTATCTGCCGGAAAACGTCGGCGCGG
CATTTATCGGCTTCGGCACCATGATGGGCAAAGACGGCAAGCCGTTCAAAACGCGCAGCG
GCGACACCGTGAAACTGGTCGATCTGCTGACCGAAGCCGTCGAGCGCGCCACCGCTTTGG
TGAAAGAAAAAAATCCCGAATTGGGTGCGGACGAAGCCGCTAAAATCGGTAAAACCGTCG
GCATCGGCGCAGTCAAATACGCCGACTTGAGCAAAAACCGCACCAGCGACTATGTGTTCG
ACTGGGATGCCATGCTCTCGTTTGAAGGCAACACCGCCCCCTATCTGCAATACGCCTACA
CCCGCGTGCAAAGCGTGTTCCGCAAAGCAGGCGAATGGGATGCAAATGCGCCAACCGTTT
TGACCGAACCGCTGGAAAAACAGCTTGCCGCCGAGCTGCTGAAATTTGAAGACGTACTGC
AAAGCGTGGCGGACACGGCGTATCCGCACTACCTCGCCGCCTACCTCTATCAAATTGCGA
CCCTGTTCAGCCGCTTCTACGAAGCCTGTCCGATACTCAAAGCCGAAGGCGCAAGCCGCA
ACAGCCGCCTGCAACTGGCAAAACTCACCGGCGACACGCTGAAACAAGGCTTGGATTTGC
TGGGCATCGATGTGTTGGACGTAATGTAAAACCGCACCGCCCGATTGCGGACAACAGCCT
CGCCATCCTTATCCGAATCTGAAAAAAGCGGCGCGATACACCGTATCCGCCGCCCCTCCC
AAAATGCGAAACAAACAAACGCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA
AGCAAGCAAAAAATTATAACCCCCTTCCTGCCGACGCACGCACTTTCCGCGCGGCGCATT
CCCCTTTTCCCGCCCCTCAAATCCGCCTTTTCTTCAGGCAGGGTTTCAGCCCGCCTCTTT
TCCCTGTTTTCCTTTCCCCGACACGCGTGCGCTCCCCCTGCCGCACTGTGCTGCACTTTC
GCGCCCGGACGGCATCGTTCCGCCATCCGGTTCTCTGTTTTACATACCCCTGTTTCAGAA
AGAAATGCAGATGTTTCAACACACAGGACGACACATAAAGCACCGCCCTATGTGTTGCCC
TGATTTGGAAGGGGTTACGCCTCCCAAATAAAGTCTGATCCTGCCGCCCCGAAGGACAGA
TGTCCGAGTGGCGAAGTTTCAACCGAAAAGGAAATACGATGAATATTCACACCCTGCTCT

CCAAACAATGGACGCTGCCGCCATTCCTGCCGAAACGGCTGCTGCTGTCCCTGCTGATAC
TGCTTGCCCCCAATGCGGTGTTTTGGGTTTTGGCACTGCTGACCGCCACCGCCCGCCCGA
TTGTCAATTTGGACTATCTTCCCGCCGCGCTGCTGATCGCCCTGCCTTGGCGTTTCGTCA
AAATTGCCGGCGTATTGGCGTTTTGGCTGGCGGTTTTGTTTGACGGGCTGATGATGGTGA
TCCAACTCTTCCCTTTTATGGATCTCATCGGCGCCATCAACCTCGTCCCCTTCATCCTGA
CCGCCCCCGCCCCTTATCAGATAATGACCGGGCTGTTGCTGCTGTATATGCTGGCGATGC
CGTTTGTGTTGCAGAAAGCCGCCGCCAAAACCGACTTCCGGCACATTGCCGTCTGCGCCG
CCGTTGTGGCGGCAGCCGGCTATTTCACCGGCCATTTGAGTTACTACGACCGGGGTCGGA
TGGCCAATATCTTCGGCGCAAACAACTTCTACTACGCCAAAAGTCAGGCGATGCTCTACA
CCGTCAGCCAGAATGCCGACTTTATTACCGCCGGCCTGGTCGATCCCGTCTTCCTCCCCT
TGGGCAATCAACAGCGTGCCGCCACGCATCTGAACGAGCCGAAATCTCAAAAAATCCTCT
TTATCGTCGCCGAATCTTGGGGGCTGCCGGCCAATCCCGAACTTCAAAACGCCACTTTTG
CCAAACTGCTGGCGCAAAAAGACCGTTTTTCGGTTTGGGAAAGCGGCAGTTTTCCCTTCA
TCGGCGCGACGGTCGAAGGCGAAATGCGCGAACTGTGTGCCTACGGCGGTTTGCGCGGGT
TCGCACTGCGCCGCGCGCCCGACGAAAAATTTGCCCGCTGCCTCCCCAACCGTTTGAAAC
AAGAAGGTTACGCCACCTTTGCGATGCACGGCGCGGGCAGTTCGCTTTACGACCGCTTCA
GCTGGTATCCGAGGGCGGGCTTTCAAGAAATCAAAACCGCCGAAAACCTGATCGGTAAAA
AAACCTGCGCCATTTTCGGCGGCGTGTGCGACAGCGAGCTGTTCGGCGAAGTGTCGGCAT
TTTTCAAAAAACACGACAAGGGACTGTTTTACTGGATGACGCTGACCAGCCACGCCGACT
ATCCCGAATCCGACATTTTCAACCACAGGCTCAAATGCACCGAATATGGCCTGCCCGCCG
AAACCGACCTCTGCCGCAATTTCAGCCTGCACACCCAATTCTTCGACCAACTGGCGGATT
TGATCCAACGCCCCGAAATGAAAGGCACGGAAGTCATCATCGTCGGCGACCATCCGCCGC
CCGTCGGCAACCTCAATGAAACCTTCCGCTACCTCAAACAGGGGCACGTCGCCTGGCTGA
ACTTCAAAATCAAATAACAACAATGCCGTCTGAACGCACCAACAGCCTTCAGACGGCATT
TTGCAGACAGACCGACCCTTCAAGCCCACTTTTTTCATCATCTCCGATAAATTGCTTTGT
ATAGTGGATTAACAAAAACCAGTACGGCGTTGCCTCGCCTTAGCTCAAAGAGAACGATTC
TCTAAGGTACTGAAGCACCAAGTGAATCGGTTCCGTACTATTTGTACTGTCTGCGGCTTC
GTCGCCTTGTCCTGATTTTTGTTAATCCGCCATAAAGACCGTCGGGCATCTGCAGCCGTC
ATTCCCGCGCAGGCGGGAATCCAGAACGTGGAATCTAAAGAAACCGTTTTACCCGATAAG
TTTCCGCACCGACAGACCTAGATTCCCGCCTGCGCGGGAATGACGGGATTTTAGGTTTCT
GATTTTGGTTTTCTGTCCTTGTGGGAATGACGGGATGTAGGTTCATAGGAATGACGTGGT
GCAGGTTTCCGTATGGATGGATTCGTCGTTCCCGCGAAAGCGGGAATCCGGAAACCCAAA
GCCACGGGAATTTATCGGAAAAACCGAAACCGCTCCGCCGTCATTCCCGCGCAGGCGGGA
ATCTAGGTCTGTCGGTGCGGAAACTTATCGGATAAAACGGTTTCTTCAGATTTTACGTTC
TGGATTCCCACTTTCGTGGGAATGACGGGATGTAGGTTCGTAGGAATGACGTGGTGCAGG
TTTCCGTATGGATGGGATTCCCTCTTGCGTGAGGCTGACAGATGCCGTCTGAAAGACTTT
CAGACGGCATAGCTTTTTCTCTTTGAATTTATAGTGGATTAACAAAAATCAGGACAAGGC
GGCGAGCCGCAGACAGTACAGATAGTACGGAACCGATTCACTCGGTGCTTCAGCACCTTA
GAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCCGTACTGGTTTTTGTTAATCCACG
ATAAATTTGCCACAAAAAAGCTGCCTCAAATGAATACCCGGGCAGCTTTTTGTTGATATG
ACTCCAATCAGCGGTGTTGCGGATTGTAACGTTTTTCCAAACGCAGGAATATCCAGCCTA
AGAAAGTCGTCATCAACAGATAAATCAGGGCGACGGTGTAAAGCGGTTCTTCATAAACCG
AATACCGGCCCGTAATCGTATTCTGAACATACGCCAACTCCGCCACAGCAATGACCGACA
GCAGCGAGCTGTCTTTCAAGAGCGTGATGAACTCGCTCGCCAAAGGCGGCAGCATGCGGC
GCAATGCCTGCGGCAGAATCACATAGCGCATCGCCTGCGGATAGGTCAGCCCCAAAGAAC
GCGCCGCCTCCATCTGTCCTTTGTCTATAGACTGGATGCCCGCGCGGAAAATCTCACAGA
TATACGCCCCCGAGTTGGCGATCAGTGCCAAAGAACCGGCAATCAGCGGCCCGTATCCGC
GACGCAGCGCGATTGCCGCCTCGCCGCTGACCAAAATGCCGTCTGAAGGATGGACGAAAA
ACGGAAACCACACATACGCCCAAATCACAATCTGCACAAACAGCGGCGTACCCCGGAACA
GCGTAACATACAGCAGCGAAACTTTACGCAACGCCCACGCCAGCACGCGCATCGGCGCAC
CGGCTTTTTCCAAGTGAATCAGGCGCGCCAACGCCAACAACAGACCCAATACCGAACCGC
CCGCCGTTGCCACGACCGTCAGCCCCAAGGTCGTCAGTGCGCCGTAAAGAAACATCCAGC
GGTATTCGTAAATAATGTCAAAACGAAAATCCATAAACCGTCCGTATCAAAAACCGGCGG
AACTGCCGCCGTTGCAAAATAATCCGCCATTTTACCGTAAAAACCGCCGCCTGAACTTTT
TTATCGCGGCAGACGGCGGTTGCGCGTCTCCGCAAAAATGCAGGGCGCGCGGTTTTCAGA
CGGCATTTGCCGTTCAAAGCCGTGCGGTGTCTTTACCAAATGCCCAACCATTCGCCCACG
GCATCCATCCAATCCTTATTGCCCCCGCCGCTCCTGCCTGCTCGGCGGTACGCCCACGGC

GCTTGCGGATTTTTAGCTTTCCACAATCCTTTGCGTTCCCTTTCCGCCTGAATTTGAGCG
TCGGCATAATCGGCAAAATCCGCCTTATCCTGCTGTTCTTTAGCATAACTTTTATAATGC
CACGCCGCCCCGTCCTGCACCTGCATCAGGTTCAAATCGGTTTTGCCGACAGAAACCTGC
GCCACTTCGCGCTGGTAGCGGTCGGTATCGAACACGCGCACGCTGACTTTCCTGCCTTCC
GCCGCCGCGCGCAGGTTGTCGCGCGAACGCGTGCCGTAAGCCTGTTTCATCTCCGGCGCG
TCGATATACGCCATCCGGATTTTGTGTTTCGCGCCGTCGCCGTCGATAACGTGAAGGGTG
TCGCCGTCATAGACTTTGGACACCGTGCCTGTGTAGCGGTGGCCGGATTTCGCCGATGCT
CGGCGGCGGGCGGGCGCGTCGGAACCCGCGTCCCCTGCCGCGCCGAGTACGTCGAGTACG
GCAACCGCCGTCCGCACCGCCTCGCTGCCGTACCCCGTATAACCCAACGCACCCAAAAGC
GACAGGGCGACGGGAAGCCATTTCATGATTTTTTTAATCTGCATATTTTTCAAATGCCGA
TGCCGTCTGAACATATCGGAATCGGATTTCAGACGGCATCTTAACGTCAGGATTACCCTT
GGCAGGGATAGATGACTTTCGCACCCTCTTCCGTCCCCAAAATCAACACATCGGCGGCAT
CGCGGGCGAATATGCCGTTTTCGAGCACGCCGGTGATTTTGTTGATTTCGTCTTCCATCG
TCAGCGGCTGATCGATATTCAAGCCGTGGACATCGACGATTTGGTTGCCGTAAAACGTGG
TGTAGCCGATACGCAGTTCGGGCTGTCCGCCCATAGCGAGCAGTTTGCGCGAAACAAGAG
AGCGCGCGCTTTCGACGACTTCCACAGGCAGAGGGAATTTGCCCAAACGTGAAACATATT
TGCTTTCATCCGCAATGCAGATGAATTTTTCGGACGCGCTGGCGACGATTTTTTCGTTGA
GGTGCGCGCCGCCACCGCCTTTAATCATTTGCAGGGCGTGGTTCACTTCATCCGCACCGT
CGATATAGACCGCCAACCCCGATACTTCGTTCAAAGAAACGACGGGAATATCGTACTGGG
CAAGCAGTTCGCCGGATTTTTTGGAAGTAGATACCGCGCCTTTGATTTTTTTGCCGCTCT
TACCCAAGGCTTCGATGAAAAAGTTGATGGTCGAGCCGGTACCGATGCCGATATATTCAT
TTTCGGGTACGAATTCGACTGCTTTTTCGGCGGCGATGCGCTTGAGTTCGTCTTGTGTCG
TCATATTTTTGTCCTTTGGGAAACCGTATCAACAAACAGCCGCCATCTTAACATTTTTTT
GCACGTCCTGCCCGCCGCGTTCAAATGCGTACCAGCAATACCGCCGCCTGCGCCTCTATG
CCTTCCATCCGCCCGAGATAGCCGAGTTTTTCGTTGGTTTTGCCTTTGATGTTGACGCAC
GAAATGTCTATGCCCAAATCGGCGGCGATGTTGGCACGCATTTGCGGAATGTGCGGCGCG
AGTTTGGGTTTCTGTGCAATCACGGTCGTATCGACATTGACCGCCTGCCAACCCTGCGCC
TGAACGCTTTGATACGCCGCACGCAAAAGGACGCGGCTGTCCGCATCTTTGAACTCTGCG
GCGGTGTCGGGGAAATGGCTGCCGATATCGCCCAAACCTGCCGCACCGAGCAGCGCGTCG
GTAACGGCGTGCAGCAGCGCATCGGCATCGGAGTGTCCGAGCAGCCCTTTTTCAAATGGG
ATTTCAACTCCGCCAAGTATCAGCTTTCTGCCTTCGGTCAGTTGGTGGACATCGTAGCCC
TGTCCGATACGGATGTTCGTCATCGTTTGTGTTCCTGATGTTTTGAATTGAAGTTCAGAC
GGCATCGAGCAGCAGCCTGACGATGTATGCGTCCTGCGGCTGCGTCAGTTTCAAATTGCG
CACGTCGCCCTGTATCAGTAGCGGACGCACACCCAATTTTTCCACGGCGGACGCTTCATC
GGTAATGCCGTCCAAGTTTTCCGCAGCCAATGCGCGGTGCAGCAGCCCGGCGCGGAAAAG
CTGCGGCGTTTGCGCCTGCCAAAGGCTCGTCCGCTCGACGGTTGCACTAATGTTCCCACC
GTCCGCGCACTTGAGCGTATCGGCAATGGGAATTGCCAAAATCCCGCCTTCGGCGGCGTT
GCCCGCCTGTTCTATCAACCGCGTCAAAGCTTCAGACGGCAGGCAGCAACGCGCGGCATC
GTGTACCAGAATATTGTCGGTTTCCGCCGCCAAACCGGTTTCCAACAGTTTTGCCACACC
GTTGCGGACGGTTTCGGCGCGGGTCTGTCCGCCGTTTTTCCACACCCGAACCTGTGGAAA
TGCCGTCTGAACCTTATCGGCAAACGTGTCTTCGGGCGAGACGACAACGACGGTCAAATC
GACGGCCTCATGCCGTTCAAAAATCCCAATCGTATGTTCTAAAACGGTTTTGCTTCCGAT
TTCGACATATTGCTTGGGTTTGTCCGCACCGAAACGCGCCCCGATGCCGGCGGCGGGAAT
CAGCGCGATATTTTTGCGCTTCATGCGTCCGTCCCGCCGTTTTCAGACGGCACGGCTTCC
TTGCGCCAGATACAGGCTTCGCCCAAGCCGTCCAAATATTGCCCGTGCGCCGCCAACTCG
TTTTCGTCCGCCCTGATGACTTTCAGTTTGCCGCTGCGTTTGGTTTCGGTATGCACCACG
GGTTTGGTTTCCATTTTTTCCTCTGCGGCCGCACCCATCAGGTCGAACTGCCGCCGCGTC
ATAGCAAGATAGACTTCGCCCAAAAGTTCGCAGTCGATCAATGCGCCGTGCAGGACGCGC
TTGCTGCGGTCGACGGAAAAACGGTTGCACAAGGCATCCAGGCTGGCTTTCTGCCCGGGG
AACATTTCGCGCGCCATCGCCAGGGTATCGGTAACGGTACAGCCGAGTTCCTCAACGGTC
GGCAACCCCATCCGGCGGAACTCCATATTGAGGAAGCCCACGTCGAATTTGGCATTGTGG
ATAATCAGTTCCGCACCGCGCAGGAAATCGGCAATCTGCCTGCCGACCTCTGCAAACGGC
GGCGCGTTTTTCCCTTCCAAAACCTGTATCGTCAAGCCGTGGACGCGTGCCGCCTCTTCG
GGCATATCGCGCTCGGGGTGGACATAGAGGTGCAGGTTTTTGTCGGTCATTTGGCGGTTG
ACCATTTCCAAACCGGCAAACTCGACCAAGCGGTCGCCGCCGTCGGCATACAGACCGGTG
GTTTCGGTATCGAGGATGATTTGGCGTGTCGTCATATCGGTGTCTTTCTTCTATCTTCGT
AAATTGCTTATTTTTTAAGCAATGTATTTTTCTGTTTTCATTTCAATGCACAAACCCACT

TATTCACAGTGTGTTCACAACATTGGGCAGGCGGATTGTGTATTTTGGGGACAATTTTTT
CAGACGGCATTCAAGGTTTTTTCCTGATTGCCGCCGCGCCTAAAAACCGCCTTTCGCGCT
TAATCAAAAATACCGACAACGGAATATTGCCCAAAGCGACAATCAGATACAACAAGGAAA
TGCTGTCAAACAAAAACAGCAACACCGCGCTCAAAACGGCAGCGGAAACCATAAAAATAC
CGTTAACGATATTGTTGGCGGCAACGGCGCGGGCGCGGAAAGTCTCGCTACTGGCGGTTT
GCAGCCAGGTATAGAGCGGAACGGAGAAAAATCCGCCGAAAAAGCCGATCAGCGTCATCA
CCGCCATCACGGGATATGCCCATCCTTGCGATAAAAACCAAAAAATGCCGTTCAGCCCTT
CAAAACGGTGTCCGTGCGTCAGCCACACCAAAACCAAGCCGCAAACCGTCAAACCCAACG
CACCAACCGTTACCCAAGCCAACATCAGGCGTTCCCTGCTGAACTTGGCACACAGTACCG
AACCGGCGGCAATACCGATGGAAAACAGAGCAAGCATCAGGTTGAAAACATTGTCGTTGC
CGCCCAGATGGATTTGGGTAAAGGTCGGCAGTTGCGTGGTATAAACCGCGCCGACAAACC
AAAACCACGAAATACCGATAATGGCGGTAAAAACGGGCTTGTGCCGCACCGTTTCACGCA
GCAGGGATTTTGTGCCACGGACAATATTCCACTCAATTTGTGTATCGGCAGCCTTGGCGG
GTACGGACGGCATAAACAGGCTGCCGACCGTGCCTCCGACGGCGACCAGCAAAACCAGTA
TCCCGACAATATAAGGCGGTACACCTGCCACCGCCGTTCCCAAAATCTGACCGAACAGGA
TGGCGACAAACGTACCCGATTCAATCAGGCTGTTGCCCATCATCAACTCTTTGTCGTCGA
GATAATCGGGCAGGATGGCGTATTTCAGCGGCCCGAACAGCGTCGATTGCGCGCCCATGC
AAAACAGACACGCCAAAAGCAGCGGGGCAGACCGGATATAAAACCCGTATGCCGCCACCG
CCATAATGATCATTTCCAGCACCTTGACCCAACGCGCCAAAACGGCCTTGTCGAATTTGT
TACCCAACTGCCCCGACAGCGAGGAAAACAGGAAATACGGCAAAATAAACAGCAACGCGC
CCAAGTTCAACATCTGTCCGGCAGGCAGGAAGCCGTTTTGCCCCAAACCGTAAAACCCAA
TCATCACAAACAGCGCGGTTTTGAACACATTGTCGTTGAACGCGCCGAGAAACTGCGTAG
CGAAAAGAGGTGCGAAACGGCGGCTTTTAACCAGTCCCAAACCGCCTTTTTTAGCGTACA
TCGTTTTCCCTCTCTTTTTCAATCAGTTTACTTGTCGAATCATCATCCATCAGGATGCGG
TGCGCCGGCCCTTCCAAGTCGTCAAACTGCCCGTTTTTGCCCGACCACCAAAAAAACCAG
CCGATGACAAACGCCAAAATAATGCTGATGGGCACCAATATAAACATGCTTTCCATCACA
TATTCCCTGTCAAATCGTTCAAAACAAAAGTCTGCCCCGACACGGTCAGATATTCGTTAC
GCAAAGTTCCGACGGGAGCTTCGTCAAAAAACAGCTCGATACGGTCTTTGACCACGCGCC
AATATTGGGGGATTTCCGTCTGACCGAACGGCGACAGGACATGATTTTCCATTCCGCCTT
CAAGTTTGACGGCAAAACGCCCGCTTTGCGGCCGTGCTTCCGATTCGTCGTCGGCAAGCA
GGATGAAAAAGCCTATATGCCGTCCCGATTGGTCATGAATACTGAAATAATGCATAAATT
TCCCACCCGCCTTTTTTCAGACGACACCAACTAAAAACAGGGCGAATGTACCAGTTTGGA
CGGGAAGAATGCAAAGAAATTCTCCCTCCCCCAGCCGAAAACACCGGCAAACCGCATATC
CCCCTTTTTTCCGTCAAAATGCCTGACTTCCGCCATTTTCACGCAAACGCCCGATTAAGC
CAAGCAATTGCAAAGATTTTTTGCTAGAATAGCCTGCTTCTTTTATCAACCTTTTCAGAC
GGCCCCACTACTTTCCCGCCCAGGAAGGCAAAACGGATTCGGCACGAATCCGGTTAGTAT
CCGTGTCCGATTCCAATGCCGTCTGAAACTTTCCGGAGTAAGAAAATGTCCCAAAAAATTG
ATCTTGGTTTTGAACTGCGGCAGCTCGTCCCTCAAAGGCGCGGTCCTGGATAACGGCAGC
GGCGAAGTCCTGCTCAGCTGCCTTGCCGAAAAACTCAACCTGCCCGATGCCTACATCACA
TTCAAAGTAAACGGCGAAAAACACAAAGTCGATCTGTCCGCACATCCCGACCACACCGGC
GCGGTCGAAGCCCTGATGGAAGAACTCAAAGCCCACGGCCTCGACAGCCGCATCGGCGCC
ATCGGCCACCGCGTCGTCAGCGGCGGCGAACTGTACAGCGAATCCATCCTCGTTGACGAC
GAAGTCATTGCCGGCATCGAAAAAATGCATCCCGCTCGCCCCCCTGCACAACCCCGCCCAC
CTCTTGGGCCTGCGTGCCGCGCAAAGCATTTTCAAAGGCCTGCCCAACGTCGTCGTATTC
GATACCTCCTTCCACCAAACCATGCCCGAAGTCGCCTACAAATACGCCGTTCCGCAGGAG
TTGTATGAAAAATACGGCCTGCGCCGTTACGGCGCGCACGGTACCAGCTACCGCTTCGTC
GCCGACGAAACCGCGCGCTTCCTCGGCAAAGACAAAAAAGACCTGCGTATGGTCATTGCC
CACTTGGGCAACGGCGCGTCCATTACCGCCGTCGCCAACGGCGAATCGCGCGACACCAGT
ATGGGCCTGACCCCGCTGGAAGGGCTGGTAATGGGTACGCGCAGCGGCGACATCGATCCT
TCCGTATTCGGCTTCCTCGCCGAAAACGCCAATATGACCATCGCCCAAATCACTGAAATG
CTGAACAAAAAATCCGGTCTGCTCGGCATTTCCGGCCTGTCCAACGACTGCCGCACCATT
GAAGAAGAAGCCGCCAAGGGGCATAAAGGCGCGAAATTGGCCTTGGATATGTTTATCTAC
CGCCTTGCCAAATACATCGGCAGTATGGCGGTTGCCGCAGGCGGTTTGGACGCACTGGTC
TTTACCGGCGGCATCGGCGAAAACTCCGACATCATCCGCGAACGCGTGATCGGCTACTTG
GGCTTCCTCGGTCTGAACATCGACCAAGAAGCCAACCTGAAAGCCCGCTTCGGCAACGCC
GGCGTGATTACCACTGCCGACAGCAAAGCCGTTGCCGTGGTCATTCCGACCAACGAAGAG
CTGATGATTGCCCACGACACTGCCCGTTTGAGCGGTCTGTAAGGTTTTATCCGCACACGA

ACTGCCTCCGGAAATGGAGGCAGTTTTTTTATCCGGCTTTCCATGCTTAAACAGCACTGC
CTCTTTTCAGACATTGACGGTTGCAGCCGCTTACCTGAACCTTATAGTGGATTAAATTTA
AATCAGTACGGCGTTGCCTCGCCTTGCCGTACTATCTGTACTGTCTGCGGCTTCGTCGCC
TTGTCCTGATTTAAATTTAATCCACTATAATGATTAACTATTTTTTAATCATGTTATTAT
TTTCCATAAAATACATGACATTAAGATGTTTTTCCACAAAAGATACACACACCGGCAAAC
ACCGGCTGTGTTTATCTTTTCTTATGCCTATTTTTTAATCATCGTATTTTTATCTTTTAA
TTTCAATACGCAAACTAACTTATACACACGGTTTTCACATCTTTAGACTGCTTCCGTGTG
TATAGTGGATATTGCCGTTTTCCTTTCTGACAAAAAATGCCGTCTGAGAACTTCAGACGGC
ATTTGAAACATCGGAATCAGCGGTTTTGTTCATACCACTCGATAAACTTGTCTGCTTTGA
CAAAACCCAGCAGCGGCTCGCTGCGGCTGCCGTCGGAGCGGACGACAAACACGCCCGGCG
GCCCGAACAGACCGTATTCTTTCAACAACGCCTGATGTTCGGGCGTGTTGGCGGTTACGT
CGATTTGGAAAAAGCGTTCCATATCGACTGCCTGATGCACTTCCGGCTGATTGAGCGTGT
AAGCCGCCATTTCTTTGCAGGAAATGCACCAGTCGGCATAAAAATCCAAAACGACGGGTT
TGTCGGGATGTTCTTTCAACGCCGTATCCATCGCTGCCTTCAGCGCGGCAGTATCGGCAA
ACATTTTGCCGTGTTCCGAAGATTTGCCTGCTTCGGCTGGTGGATTGAGGGTCAGGAAAT
GGTGCAGCGCGGTCGTTTTGCCGTTTGCGCCCTGCCAGCCGAACCACGCGCCGCCTATCA
GCAATATACCGCCCAATGCGAATGCCACAGCTTTCGGACGGCGTTTCTGCCTGCGTCCGT
TGACCAGCAGCATAAAGGCAGGAACCAGCATCAGCAGCGTGTACAGCGCGACGACGAGAT
AATAGGGCAAGTGCGGCGTGGCGAGGTAAACGGCGACGGCTAGCAGGATGAAGCCGAATG
CGTATTTGACGGCATTCATCCAATCGCCTGCCTTAGGCAGGATATGCCCGCCGAACGTGC
CGATGGCAATCAGCGGAACGCCGGTGCCCAACGCCAAAGTGTAAAGTGCCAAACCGCCTA
AAACCGCATCGCCCGTCTGACCGATGTAGCCCAAAGCAAATGCCAGCGGCGGGGCGACGC
ACGGCCCGACAATCAGCGCGGACAATATGCCCATAATAAAGACGGAAACGATTTTACCGC
CTGAAAGCCTGCTGCTTTGATTCTGAAAATACGACTGCACGGCGTTGGGAAGCTGGATGT
TGAACAGCCCGAACATAGACAGTGCCAAGACGACCATTAAAGCCGATGCCGCCAATACCA
CCCAAGCCTGCTGCAACCATACGGTCAGCAGTGCGCCCGTCAGTCCGGCAACAATGCCGA
CCAGCGTATAAGTCAGAGCCAAACCCTGAACATAAACGACGGACAGCACAAACGCCCGCG
CCTTGCCCGCCTTTTTGTCGCCGACCACAATACTGGAAACAATCGGCAACAGGGGATACA
TACAGGCGGTAAAACTCAGGCCCAAACCAGCGAGAAAAAAACGCCAAAAGATTGGCGTTGA
GCGTATCCCAAGACAGCTTGAAACGGCTGTCGCCGCCCTCATCCCCCTTCGGGGGCGGCA
GCGCCCCGCTGCCGTTTTGAGAGGAAGGCTGCAAAAAGCGGTCTTTGGCGGATGCCGGTT
CGTCGGTTTGCGGATGGTAAGTGCCGTTGCCGAAAATATCAAACTCGGTATCCACGGGCG
GATAGCACACGCCGGCTTCGGCACAGCCCTGATAGGTCAAAACCAATTTATACGGTTCGC
CGACAGCCTTTGCATAAGGAAAGGCAACCTGCGCCTCGTGATGGTAAACCGTCTGCCTGC
CGAAAAACTCGTCTTCCTTCTCTTCGCCCTTGCTGAAAGAAGGCTGTCCCAACAAATCCG
CCGGATCGGTCTTGCCGACGATTTTCGCCTGATACATATAGTATCCGTCGGCAATCCTGA
AACGGACGTTCACACCGTCGTCGGCAACGGCAAGCTCCGGCACGAATGCCTTTTCCGGCG
GCAGCAGATCGTTCGCATCCAGCGCGAAAGCTCGTCCGCACAACATCAAAAATACGGCGA
ACAGGCAAATCAGTTTTTTCATAATCGAATCCGTTTCAGACAAATAATTTGTCTGCATTA
TAAATGGTAAGGTTGACGGTGGGATTTAATTTATGTAAAACCCGCCATTATCCGAACCTA
TTTCCATAAACATCTTATCGAACCCGCCATGTACGATGTCAATACCCACGATGTCCGCCG
CTTTTTCGCCCGCGTGTGGCAGCAGCGGCTCAATCCGCTGCAACTGAGCGCACTGGAACA
GAAAGCCCTCCGCATTGTCGAAGCCCATCCCGAATACCACCGTTATCTCGAACGCATCGA
AGACCATCTGGACACCGACTGGCTGCCCGAAAACGGCGAAAGCAACCCCTTCCTGCATAT
GTCGCTGCATCTGTCCGTCCAAGAACAGGCGGGCATAGACCAGCCGCACGGCATACGCGC
AATCCACGACACCCTGTGCGCCAAACGCGGCTGGCTGGAAGCCGAACACGAAATGATGGA
GGCACTGGCGGAAACACTGTGGACGGCGCAACGCTACGGCACCGGTTTGGATGTCAATTT
CTACATGACCCGACTGCGCAAACTCATCGGCTTGGGTGCAGAAGATCAAGCCAGATTGAA
CCCGCATGAAATCGCCTGACCATACCAACCGCCTGCAAAATGCCGTCTGAAGCGGAACAA
CCCCTTTCAGACGGCATTCATTTTCCCCCAATCATTTCCACAACGCCTTTTTCAGCATAA
TCAACCAATCCTTCTTATCCAAAACGGGGCGTTGTGCAAACACATCGTATCGGCACGCGT
CCAGTTTCTGCAAAATCAACTGCGCCCCCAACACAATCATACGGAGTTCCAAACCGATAC
GCCCATTCAGTTCCCTTGCCAAAGGCGAACCCGCCTTCAGCATACGGAACGCACGCCGAC
ACTCATACGCCATCAGCCGCTGAAACGCCGCATCCGCCCGTCCTGCCGCGATCTGTTCCT
CAGAAACACCGAATTTCAACAAATCGTCCTGCGGAATATAAACCCTGCCTTTTTGCCAAT
CCACAGCCACATCCTGCCAAAAATTCACCAGTTGCAAAGCCGTACAGATGCCGTCGCTTT
GCGCCACGCACACCGCATCCGTTTTCCCGTACAAAGCCAGCATAATGCGTCCGACAGGGT

TGGCGGAACGCCGACAATAATCGGCCAGCTCGCCGAAATTTCCATACCTTGTTTTAACCA
CATCCTGAGAAAATGCAGAAAGCAAATCATAAAACGGCTGCAAATCCAAACCGAACGGCA
CAACCGCCTCGGCATCCAATCGTGCAATCAAAGGATGCGCCGACCGGCCGCCCGATGCCA
ACACGTCCAACTCGCGCTGCAAACCCTCCAACCCCGCCAACCTGGCTTCAGACGGCATAC
TGCCCTCGTCCGCCATATCGTCCGCCGTCCGTGCAAACGCGTACACCGCGTGAACCGGCT
TCCTCAACCTGCGCGGCAAAATCAGCGAACCGACGGAAAATTCTCATAATGCCCAACCG
ACATACCTTCTCCATCCATCAAACAAAATGCCGTCTGAAACGGAACAAACCCTTTTCAGA
CGGCATCAGATACCTCCAAGCTGCCGGCAATCAGTGGTGGTGATGACCGTGCGGGCCGTG
GACATGACCGTGTGCGATTTCCTCATCGGATGCATCGCGCACGCTTTCAACTGTAGCCTT
AAAGCGGATTTTCATGCCTGCCAAAGGATGGTTGCCGTCCACCACCGCCTTGCCGTCGGC
AACATCGGTTACACGATAGACGACAACATCGCCGGTTTCAGGATCGTCGGCTTCAAACAT
CATGCCGACTTCGACTTCAACAGGGAACACGCCCGCATCTTCGATACGGACCAACTCCGG
ATCCTGCTCGCCGAACGCATCGTCGGGCGACAGCGCCACATCGACCGTATCGCCGGCATC
CTTACCGTGCAACGCCTCTTCCACCAAAGGGAAAATGCCGTCGTAACCGCCGTGCAGATA
CGCAATCGGTTCTTCGGTTTTGTCCAAAAGCTGATTGTTGGCATCATACATCTCATAATG
CAGCGAAACCACGGAATTTTTCACGATAGCCATATTTGTCCTTTCAGGAACAGCAGATTA
ATTACAGGCGCATTCTAACACAACCGCCGCGCCGGCCGATTACCGTTAACCTGTTCATAA
ACTGTACAGCACATATTTCAATGTAAATCTTTGTTATTTTATTGCGGTGTAACTTTTTTA
CAACATTCTTAAAACCATTCCGACCTGTCTGCCGACTTTCCCAATCCGCCTTAATAAATC
ATACAAGATACTGAAATTATATTAATCTCTATAATATTTATCCCTATCGAATTTTTAACA
GCAAAACCGTTTTACAGGATTTATCAATCCGCCCGCCAGAAAACTTTTCATTCAAACCTT
TTTCCCATCTGTACGACATTGCAATCCCTTATTCCATAGTGCATAATTACGCAAATTCAG
CGATGAATTTCCAACCCGGTTTGTAGTATGGTCGATAAAGACCTATTTGTTTCAATAATT
TAAATTGGTTCTAAAGGTTACTAAAATGAAAAAATCCCTGTTTGCCGCTGCTTTGTTGTC
TTTGGTTCTGGCAGCCTGCGGCGGTGAAAAAGCCGCTGAAGCTCCCGCTGCTGAAGCACC
TGCCGCCGAAGCTCCCGCTACTGAAGCACCTGCCGCCGAAGCTCCCGCTGCTGAAGCACC
TGCCGCCGAAGCTCCTGCTGCTGAAGCTGCCGCTACCGAAGCACCTGCCGCTGAAGCTGC
CGCTACCGAAGCACCTGCCGCTGAAGCTGCCGCTACCGAAGCACCTGCCGCTGAAGCTCC
TGCTGCCGAAGCTGCAAAATAAGCATTTTCCGCTTGCAAAAAAGCAGGATACGTTCAGTA
TCCTGCTTTTTTGATTTTTCAGACGGCATCAGATTCCCTTCCTCAATCTTCTCCCTACCC
TTCCGACAAACATGCTTGACCTTCATACCGAATTTTCCCGACTCCTACCGGCAGATGAAA
TTGCCGAACCTTCTCCGACGCTTTTAAAAGACCAGCGCAACCGCTTTACGTCTGCACCAG
ACATCATTTTGCAGCCGCTCAGCGTTAAAAGCGTGCAAACCATTATGCGTTTCTGCCACC
AACACCGTATTCCGGTTACGCCGCAAGGCGGCAATACTGGTTTGTGCGGCGCGGCAGTAT
CGGAAAACGGCGTATTGCTGAACCTTTCCAAACTCAACCGCATCCGCAGCATCAATTTGT
CAGACAACTGCATAACCGTCGAAGCAGGTTCCGTACTCCAAACCGTCCAACAGGCAGCCG
AAGCCTCAAACAGGCTGTTCCCACTCAGTCTCGCCAGCGAAGGCTCGTGCCAAATCGGCG
GCAACATCGCCTGCAATGCCGGAGGTTTGAACGTATTGCGTTACGGCACGATGCGCGACC
TGGTTATCGGTTTGGAAGTCGTCCTCCCCAACGGCGAACTGGTTTCCCATCTCCATCCCC
TGCATAAAAACACCACCGGCTACGACCTGCGCCATCTGTTTATCGGTAGCGAAGGTACAT
TGGGCATTATCACTGCCGCCACGCTCAAGCTGTTTGCCAACCCCTTAGACAAAGCAACCG
CATGGGTCGGCATACCCGACATCGAATCCGCCGTCCGCCTGCTGACCGAAACCCAAGCAC
ACTTTGCCGAACGCCTATGCAGTTTTGAGCTGATCGGCCGTTTTGCCGCCGAATTGTCTT
CCGAATTCAGCAAACTCCCCCTGCCGACACATTCAGAATGGCATATTTTACTTGAGTTGA
CCGACTCATTACCCGACAGCAATCTTGATGATCGGCTTGTCGAATTTCTTTATAAAAAAG
GCTTTACCGACAGCGTGTTGGCGCAAAGCGAACAAGAACGTATCCATATGTGGGCGTTGC
GCGAAAACATCTCCGCATCGCAACGCAAACTGGGCACCAGCATCAAACACGATATTGCCG
TTCCTATCGGGCGCGTTGCCGACTTTGTCCGCCGGTGCGCCAAAGATTTGGAACAGAATT
TCAAAGGCATACAAATCGTCTGCTTCGGACATCTGGGCGACGGCAGCCTGCACTACAATA
CTTTCCTGCCCGAAATCCTCAGCAATGAAGTCTATCGTTACGAAAACGACATCAACAGCA
CAGTCTATCGCAACGTCCTTGCCTGCAACGGCACGATTGCCGCCGAACACGGCATAGGTA
TCATCAAAAAACAGTGGCTGGACAAAGTACGCACGCCTGCCGAAATCGCCCTGATGAAAA
GCATCAAACAACACCTTGATCCATATAACATTATGAATCCGGGCAAACTGCTTCCGTAAC
CGGCATTTCTGATTTGCATACACAACAAAGAAAGGGACAATAGATCCGATTGTCGGTTTA
GCGCGAGCTCGTGAGTGCGGTTAAAAATTGGTGGAAATTACACGAAAAATGACCGCACTT
TTAAAATAAAAAAATCGGCAGTGAATTTCCCTGCCGATTTTATTTTGTTACAACTTAACT
TAAAACGTCCACTGTAAATTCAACGCACCTTGTTTAGCTTGATGATGTTTGCCTGTTTGG

CGGTTGAATGTGGCTTGTAAGGTTAAGTGAGATTTGATTTTCACTGCTACACCTAATTGG
CTCTCAATTGCCGTCTTATTGTTTATCACTCGACGCTCTCCGTCCATTTCCACACCGAAA
GGTTTGTTGTGGTAAAGCGCGTTCACAGCGGCGAAAGGTTCAATAGCGATATTTTTATAG
AGTGAAAATTGAGCTTTAGCTTGAACGCCAACCCGAGTTTGTAATTGGCGGGAGCCAAGT
AAATTCACGTGGGCATTTTCGCTATCGCTGAATTTTCCGTTTACCCCCAAATAAGTCAAT
TGTGCCTGTGGTTGTAGGTAAACACGAAGGCTGTTGCCCTTTTTAGTGAAGTGTTCCGCC
AATAACGCATTGTAACCTGCTTCAATTGAGGCAGTAATACCTTTTGAAGTAAAACGTTCT
GTACCATCTTCAGTGTTGATACGGTGGCGGAAGCGTTGATATTGCATCCAGCTATCCGCA
TACGCACCTGTCTGTTTGTCCTGAAGTTGGTGCCAAGTGGCGTAAACGCCTGCACCAAAG
CCTTTCACATTTCCCGTTGTAAGATTGTCTGTATCTGGGTTGTGGAAAGTGCTACGTTGT
TCTGCTTGTCCGCCCATTAAGCCAATAGAAAGTTGATTACTTTCGTTTTGCCATGTGAAT
ACTTCGCCGCCGAGTTGCACACCTTTACGATAGCCTTCTACAGGTGCTGTTTTGCCTTGC
ACCCATTGGTTGGAATGTCCGTCAATCACACGCAACCACAAGCCTTTGCGTGGTAAAGTG
CGGTCGAAAATATCGCTGTTTTTGTTGTTCAAACGCAAGGCGAATAAGGTATTGGCGGCT
TGAGCCTGTTGTGCATAAATCGCCATATCATCGCGTTCTTGCACTTTGGTAAAAAAGCCC
TCTGGGCGTTGTTGTAAAGAAAGCGTATAAATTCCCTTTTGGTGTTTGCCAGAAAGACGG
AATGCGTGTTTATCTGCTGTGCCATTTACTTTGATAATTTGATGCCCATCGAGGCTTTTT
AAATCGTCTATTGGATTTTCGAAGATGATGTCGGAAGTGCCAGTAACATTTTTCTCAAAA
ATTAATGCAGTATTTTTCGCTTCTTTAGGATCGTAAGCAAAACGAAAACGAGCTCCGCCA
GCATAATCTTCTTTTACGAGTAAACTTTCACTTTTAGTATTAAAACGGATGTCTGCATTC
GTTGTTTTTAATTTCCCAACATTAGAATCCCAACGGGGCTCCCAGAGAGAATTTTCTAAG
CGGAATTCATCCAAACTAATCGTTTGCCCGATAACGTGCGAGTTGTCTGTAACCTCAATA
TAGTGGAATGGATCTAAACCAGAATATAGATGTGCTGCAAAAGAAACATAATTTTCAATA
TGATGAATTACTTGATTAGCCCATTCTGTATAATTCCCGACAGATAAAATTTCGCTGTTG
ATATGACTATTTTTTATTTTTGGACCTAAGGAGAATATATGACTTTTTACTATAAGAGGA
TGGGATCCAAATTTTTCAGCTTGGCAAGTACTATAATCACGTATCTTAGTGTTAGAATTA
AAACATTCCTTAAAATATTTCCGTATTTGTTCTTCTGTGTCCCCATTTCTTTTTGCAACC
CCTAAACCTCGGGCGAAGCCAACTAGGTAACCTTCGGTATATTCTTGATCATAAAAAGAA
ATCTTTTTTGAGTTATTGATGTTTTCGAATTGGTATGTTCTAGGGTATAGTGCGGGAAAG
GGTGGAACTTTTGGATTATCCTCGGTTATAAGATAAGTTTCTTTTTTCCAATATTCACTC
GTTTTATCGCGGAGTTTTTTTAAGCGGGTAATTTCATCATTAGTGAGCTTGGTTTTGTCG
TAAACGTAATCAACAGCCAAAAGCGGAGAGGTATAAAGAATAGAAAAAAATAGACTTACA
ATAAATGATTTTTTAAACTTCTGCTTGCTTGCTTGCTTGCTTCGAGTTTCATAATAAATT
TTCCTTTGTCAAGTAAAAATAAATGGGGCGTGGATTTAGCATAAAACTGAACAAAAAAT
GTCATTTATCTCACATTTTTCTCTATTTATTTCTTGTTTATTAAAAGTAAACGTTTGCTT
TTTGCTATTTTGTCAAGCCAGTTTGAAAATGTGTATAATTGCCCTCGTTATTTACAAAAA
TTTCAGGAAAAATGACCGCACTTTACCCTTGGCTAATGCCAATTTATCATCAAATTGCTC
AAACCTTTGACGAAAGCTTGGGGCATCATGCCGTGCTGATTAAAGCGGATGCTGGTTTAG
GTGTAGAACGTTTACACATCAGGCGGCAGCCTTGCCCATACCGTCTGAAGCACTGTTTCC
ACAATCAGCGCGTATGCTTAATCAACCGCTGTTTCTCGCGTTTCCAATCCGCCTCTTTCA
TACTCTGGCGTTTGTCGTGCTGTTTCTTACCTTTTGCCAAACCGATTTCCATCTTGATTT
TTCCGCGTGAAAAATGCAAATCCAGCGGCACGATGGTGTAGCCGGCACGTTCGGTTTTGC
CGATTAATTTGTTGATTTCCGACTGGTTCAACAAGAGCTTGCGCGGACGTACGGCATCTG
GTTTAATGTGTGTCGAGGCTGTGGGCAAAGCCGTAATATGGCAGCCGACCAGATAAAACG
CGTCTTTTTTCCAATAGATATAACTCTCTTTAAGCTGTACGCGCGCGGCGCGGATTGCTT
TGACTTCCCAGCCTTCCAAGACCAAACCGGCTTCAATCCGGTCTTCAATGAAAAAATCGT
GAAATGCTTTTTTATTGTTCGCAATAGCCATAAACATCCTATCAATATCCGCCGTCAGAC
GGCATAAACCCGAAAACAGAACCCATCATACCGCCTCTTCAACCGCCTGCACAATCTTCT
CGGGATACAGCCTGTTGAGGCAGTCGGTATGCCCCAGCGGACATTCCCGCTTAAAACACG
GCGAACATTCCAAGTGCAGGCTGACGATTTTCGCCCTATCGCTCAAAGGCGGCGTATGCG
TCGGGCTGGAAGAACCGTAAACCGCCACCACCTTCCTGCCCAAAGCTGCCGCCAAATGCA
TCAATCCGCTGTCGTTACACACGACCGTGTCCGCCAACGACAGCAAATCCATTGCCTGCG
ACAAATCGGTTTTGCCGCACAAATTGACACACATACCGTCTGAAAGGCGGTTGATTTCCT
CGGCAATTTCATCATCTTTTTGCGAACCGAACAGCCAAACCTGCCAACCCGCCGCCAGAT
AATGTTTGCCCAACTCGGCAAAATGCCTTGTCGGCCAACGCTTTGCCGGCCCGAATTCCG
CACCCGGACAAAAAGCCAGAACAGGCTTTCCAATATCCAAGCCAAAGGTTTCGACAGAAA
TTTCCCGCCGCCGTTCATCAATGGAAAACTCGGGGAATCCCGAATGCCCGTCAAAATCTT

CCTGACTCGGATGCGCGAGAGCCGTATATCGATCCACCATCAAAGGCAGACGTTCCTTAT
CCAGCCTGCGTATATCGTTCAACAGAAAATAACGGCTTTCACCGACATAACCCGTCCTTT
TACCGATACCTGTCGCCAGCGCGATGATTGCCGATTTCAAAGAACCGGGCAACACGATAA
CCTGATCGTATCCGCGCCGCCCCAAATCCCTACCGACCCGCCAACGGCGTTTCAACTCCA
ACGCACCATGTCCGAACGAATTCTCAAGAATTTCATTCACTTCCGGCATACGCTCGAACA
CCGCCATCGACCACTTCGGTGCGAACACATCAATCGTGCAACCGGGGTGAAGTTCCTTCA
AACGGCGGAACAAGGGCTGGGTCATCACGCAGTCGCCTATCCAACTGGGGGAAATAATCA
GGATTTTGATGGACATAACAAGAAACCGAAATCAGACAGGCAGAATTTTACCGCGAAACC
GTTGGAAAACCTATCTTGCCGCATTCCGAACGCCGGACGTGCAAATATGAAAAAGCCCGA
ACATTCAAGTTCGGGCTTCAAAATTCTGGCTCCCCGACCTGGGCTCGAACCAGGGACCTG
CGGATTAACAGTCCGTCGCTCTACCGACTGAGCTATCGGGGAATGGGGCGTATTATAGCG
TC

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 3>:

gnm_3

GCGGGGGCtTCCATCGCAGTCATGCACAACATCTGCATCAAATGGTTTTGCACCATATCG
CGCAACGCGCCGGTAATGTCGTAAAACTCACCGCGCTCTTCCACACCGAGCTGTTCGGCG
ATGGTCAACTGCACGCTTTCGATATATTTATTGTTCCACAGCGGCTCGAACATTACATTG
GCAAAACGCAGCGCAAGCAGGTTTTGCAGGCTTTCTTTGCCAAGGTAGTGGTCGATGCGG
TAAATTTGCCCTTCTTTGAAATAACGCGCAACATCGGTATTGATTTGCTGGGAAGAAGCC
AAATCCGTACCCAACGGTTTTTCCAAAACTACGCGCACATTGTCGGCATTCAAACCGATC
GCAGCAAGGTTTTCACAGGCTTGCGCGAAGAATTTGGGCGCGGTGGACAGATAGATGACG
ACGTTGTCGGTTTCTTTGCGCGCTTTGACCAAATCGCCCAAAGCGGCAAAATCGTCCGGC
TGCGTAACATCGACTTTGAGATATGCGAAACGTTCGACAAACGATGCCCAAGCCTCATCG
GAAAAATTTTCTTTCACATGGATTTTGGAACTGGTTTCCACCTTCGCCAGAAAACCTTCG
GTATCCAACTCGCTGCGGCTGACCCCCAAAATACGCCCTTCGGGATGAAGCAGACCGGCA
ACATGCGCCTGGTACAGACAGGGCAACAGCTTGCGCATCGCCAAATCGCCGGTCGCACCG
AACAACACCAAATCAAAATTTGTTTGTGTACTCATCGTATTATCTCGTCAGGAAAGAATT
TTTCGATGCCGTCTGAAACCTGTTTCCCCCATCACGCTGCATCGCAATATCGGAAACAAA
GGCAGGCGGCATAATGAGTAGTAATACTACACACCGCTACACTTTTTGTCTATTCCCATT
TTTACAATTTATTTGACCTAGTCCAAAAATCGGGCAGGTTTCCCCTATTCCGTTACAACA
ATCGAAAGATTCTGCGATTTAAATCAAATTTCTTTTCAATGCCTGATTTTTTTGTAACAA
AATTACAAATTTTGTACTATAATAACACCCGCTTCCCACTTTCAGACGGCATACCTTTTA
AAATATAGTGGATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAGATAA
TACGGAACCGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGG
CGAGGCAACGCCGTACTGGTTTTTGTTAATCCACTATACTTACCGTCTGAATACCCGATA
CAAAAATCAGAAACGCACAAACAAATCCCCAATACCCCCCCCGTTCCGACAGGAGACCGA
CCGTGAACACTACTCCTATCCACTCCAAACTCGCCGAAATCACCGGGCGCATTATTGAAC
GCAGCCGTCCGACGCGTGAAAAATATCTGGCGAAGATCCGCAGTGCCAAACAAATGGGAC
GCTTAGAGCGCAACCAGCTCGGCTGCAGCAACTTGGCACACGGCTATGCTGCCATGCCTA
AAAGTATCAAAATCGAAATGCTTCAGGAAACCGTCCCCAACTTAGGCATCATCACCGCCT
ACAACGACATGGTTTCCGCACACCAGCCGTTTAAAGACTTCCCTGACCAAATCAAAGACG
AAGCGCAGAAAAACGGCGCGACCGCCCAAGTCGCCGGCGGCACGCCCGCCATGTGCGACG
GCATCACGCAAGGCTACGCCGGCATGGAATTGTCGCTGTTCTCCCGCGACGTGATTGCCA
TGAGTACCGCCATCGGGCTGTCGCATCAAATGTTTGACGGCAGCCTGTTTATGGGCGTAT
GCGACAAAATCGTTCCAGGTTTGATGATAGGCGCGCTTTCGTTCGGTCATATTCCGGGTA
TCTTCGTCCCCGCAGGCCCGATGTCCAGCGGTATCGGCAACAAAGAAAAAGCCCGCACCC
GCCAGCTTTTCGCCGAAGGCAAGGTCGGACGCAACGAACTTTTGAAAAGCGAAATGGGTT
CTTACCACAGCCCGGGCACCTGCACTTTCTACGGCACGGCAAACTCCAACCAAATGATGA
TGGAAATGATGGGCGTGCACCTGCCTGCCGCCGCCTTCGTCCACCCTTACACCGACCTGC
GCGAAGCGCTGACCCGCTACGCCGCCGGACACCTCGCGCGCGGCATCAAAAACGGCACGA
TTAAACCTTTGGGCGAAATGTTGACCGAAAAATCCTTTATCAACGCCTTGATTGGCCTGA

TGGCAACCGGCGGTTCGACCAACCACACCATGCACCTCGTCGCTATGGCGCGTGCGGCCG
GCGTGATTTTGAACTGGGACGACTTCGACGAAATTTCCTCCATCATCCCGCTGCTCATCC
GCGTTTATCCGAACGGCAAGGCCGACGTGAATCACTTTACCGCAGCGGGCGGACTGCCTT
TCGTTATCCGCGAATTGCTGAATGCAGGCCTGTTGCACGACGATGTCGATACCGTCGTCG
GACACGGTATGCGCCACTACACCAAAGAGCCTTTCCTTATCGACGGCAAACTCGAATGGC
GCGAAGCCCCCGAAACCAGCGGCAACGACGACATCCTGCGCAAAGCTGACAACCCGTTCT
CCCCCGACGGCGGTCTGCGCCTGATGAAAGGCAACATCGGACGCGGCGTGATTAAAGTGT
CCGCCGTGCGCGAAGGCTGCCGCATTATTGAAGCGCCTGCCATCGTGTTCAACGACCAAC
GCGAAGTGTTGGCTGCGTTTGAACGCGGCGAGTTGGAACGCGATTTTGTGTGCGTCGTCC
GCTACCAAGGCCCGCGTGCCAACGGTATGCCCGAATTGCACAAACTGACCCCGCCTTTGG
GCATCCTGCAAGACCGCGGCTTCAAAGTGGCGCTGCTGACCGACGGCCGTATGTCCGGCG
CGTCCGGCAAAGTTCCAGCCTCCATCCACATGACACCCGAAGCCCTGATGGGCGGCAACA
TCGCCAAAATCCGTACCGGCGACCTGATCCGCTTCGACTCCGTTAGCGGCGAACTCAACG
TCCTGATTAACGAAACCGAATGGAATGCCCGCGAAGTCGAAAGCATCGACTTGGGCGCGA
ACCAACAAGGCTGCGGCCGCGAACTCTTCGCCAACTTCCGCAGCATGACCAGCAGCGCGG
AAACCGGTGCCATGAGTTTCGGCGGCGAATTTGCCTGATGCGCGTTTCAGACGGCCTTTT
CAGACCGAAGGCCGTCTGAAAAATTATTCAAGCGTTTTAAGATAGACGTAGGTTGGATTC
TCGAATCCGACACAGCCGTCCAAGATGTCGGTTTCTTGAATCCGACCTACAACCTGTCCC
ATCTTAATAAAATACCCCATTCCACCCGGAGAACCGAAATGTCCAAACTGACCCCCCGCG
AAATTTTGACCGCCGGCGCAGTTGTGCCGGTAATGGCGATTGACGACTTAAGCACCGCCA
TCGATTTGTCCCACGCCCTTGTCGAAGGCGGCATCCCTACCCTCGAAATCACCCTGCGCA
CCCCTGTCGGCCTCGATGCCATCCGCCTGATTGCCAAAGAAGTGCCCAACGCCATCGTCG
GCGCAGGTACGGTAACCAATCCCGAACAGCTCAAAGCCGTCGAAGACGCAGGCGCGGTTT
TCGCCATCAGCCCGGGGCTGCATGAATCCCTCGCCAAAGCCGGCCACAACAGCGGCATCC
CCCTGATTCCCGGTGTTGCCACCCCGGGCGAAATCCAACTGGCTTTGGAACACGGCATCG
ACACCCTCAAACTCTTCCCCGCCGAAGTCGTCGGCGGCAAAGCCATGCTCAAAGCCCTGT
ACGGCCCTTACGCCGATGTTCGCTTCTGCCCGACAGGCGGCATCAGCCTCGCCACCGCGC
CCGAGTACTTGGCACTGCCCAACGTCCTGTGCGTCGGCGGCTCTTGGCTGACACCGAAAG
AAGCCGTGAAAAACAAAGACTGGGACACCATCACCCGCCTCGCCAAAGAAGCGGCGGCGT
TGAAACCCAAAGCCTGATTCGCATCGTAAAAATGCCGTCTGAAAAACCTTTCCCGTTTCA
GACGGCATTTTGCCGATTGAGGGCACAGTCGGCATACACGGCAGCACTGATCAGACATAC
CGCCCCTAAAATGCCCATCCGCCTTCCGCATAATAAAAATAACGTTCAGTTCATTCGACA
GCAGCCGGACAGCCCATACTACGCGGCTGAAAAAATGCCGTCTGAAACGCATTCAGACGG
CATCCACTTAAAAAAAACAACTGATTCAACGCCGATTAATCCGCTTCCAAAACCACTTTC
ATCACTTGGTTTTCGGCGGCGTGTTTGAACACGTCGTAGGCTTTTTCCAATTCACTGAAT
TTGAAATGATGGGTCAGCATTTTGGTGTAATCGACGGAGCTGCTGGAAATCGCCTTCATC
AGCATTTCGGTGGTATTGGCGTTTACCAGACCGGTAGTGATGGCAAGCTTTTTAATCCAG
AGTTTTTCCAGTTTGAAATCAACGGATTGACCATGTACACCAACCACAGCGATATGGCCG
CCGGGTTTCACAATGTCTTGGCACATATTCCATGTAGCAGGGATACCGACGGCTTCGATG
GCGCAATCCACGCCGTCTTCGCCGACGATGGCAAAGACTTGTTTGGATACTTCGCCGGAA
GCAGGGTTAATGGTATGGGTCGCACCCAATTCTTTCGCCAGTTTCAAACGGTTTTCGTCC
ATATCGCAAACGATGATGGCGGCGGGACTGTACAGTTGGGCGGTCAACAGGGCGGACATA
CCGACAGGGCCTGCCCCAGCGATGAATACGGTGTCGCCGGGTTTGACATCGCCGTATTGC
ACGCCGATTTCGTGGGCGGTCGGCAAAGCGTCGCTCAACAACAGGGCGATTTCTTCGTTG
ACATTATCGGGCAGCGGAACGAGGCTGTTGTCGGCATAAGGCGTACGGACGTATTCGGCC
TGAGTACCGTCAATCATGTAACCCAAAATCCAACCGCCGTTACGGCAGTGTGAATAGAGT
TGGGTTTTGCAGTTGTCGCAAGTGCAACATTTGCTGACGCATGAAATAATGACTTTATCG
CCGACTTTGATGTTTTTTACAGCCTCGCCGACTTCTTCTACAATACCGATGCCCTCATGA
CCGAGAATACGGCCGTCGGCAACTTCGGGGTTTTTGCCTTTCCAAATACCCAAATCGGTA
CCGCAAATCGTGGTTTTGACGATTTTCACCACCGCATCGGTCGGATCGATAATCTGCGGA
CGGGGTTTTTCTTCAAAACGGATGTCGTTTGCGCCGTGATAAACCATTGCTTTCATGCTG
ATACTCCTTGCTTGTTGATAAATAATTTCAATACCGCAATAAAGTTTCTTTATATGAGTT
ATATGCCCCTACAAAAAATAAGTCAATAAGAATTATTTTCACAATGTTATACAATAACAT
ACCGTTTTAAATATAAATAAAACCACCGATTGATATTAATGAACACACCCATCCCCTTCT
CCGAACGGCTCATCCGCTGGCAAAAACAACACGGTCGCCACCACCTCCCTTGGCAGGTCA
AAAACCCTTATTGCGTCTGGCTTTCCGAAATCATGCTCCAGCAAACGCAAGTCGCCACCG
TGTTGGACTACTATCCGCGCTTCTTAGAAAAATTCCCGACCGTTCAGACGCTTGCCGCCG

CGCCGCAAGACGAAGTGTTGTCGTTGTGGGCGGGCTTGGGCTATTACAGCCGCGCGCGCA
ACCTGCACAAAGCCGCGCAACAAGTCGTCAGGCAATTCGGCGGCACGTTTCCGTCGGAGC
GCAAAGACTTGGAAACCCTCTGCGGCGTAGGCAGAAGCACCGCCGCCGCCATTTGCGCCT
TCTCCTTCAACCGCCGCGAAACCATTTTGGACGGCAACGTCAAACGCGTACTCTGCCGCG
TGTTCGCCCGCGACGGCAATCCGCAGGACAAAAAATTTGAAAACTCGCTCTGGACACTTG
CCGAAAGCCTGCTGCCGTCTGAAAACGCCGATATGCCTGCCTATACACAAGGTTTGATGG
ATTTGGGCGCGACCGTGTGCAAACGGACGAAACCCTTGTGCCACCAATGCCCGATGGCGG
ACATCTGCGAAGCGAAAAAGCAAAACCGCACCGCCGAGCTGCCGCGCAAAAAAACCGCCG
CCGAAGTACCGACCCTGCCGCTTTACTGGCTGATTGTCCGCAACCGGGACGGCGCGATTT
TGCTGGAAAAACGCCCCGCCAAAGGCATTTGGGGCGGGCTGTATTGCGTGCCGTGTTTTG
AAAGTTTGAACGGGCTTTCCGACTTTGCCGCCAAATTCTCCCTGACCATGGCAGATATGG
ACGAACAAACCGCCCTGACCCACCGCCTGACGCACCGGCTGCTATTGATTACGCCCTTTG
AAGCACAAATGCCGTCTGAAAGCCCTTCAGACGGCATTTGGATAAAGCCGGCGCATTTGA
AAGATTACGGTTTGCCCAAGCCTTTGGAAATTTATTTAAACGGTAATAGGTTAGAATAAA
CAAAATAAACCCATTGAACTGTTGTTTGCAGGTATCGCAGCAAGAACAACCGATGAATTT
GGGTCGTATTTTAGGCGGCGGGATAATGTTCAAATGGGACATTTGGAACGGAAGAAGTCG
GCAATTTAAAAAGGATTTAAAAAGCAAAGAAGGTCAAAAACATGAACACAAACTTAAATG
ACAAAGACAAAGCCATGGATACCGCAATCAGGTTTCAGAAAAGGATGAGGATTCCGAAAT
TTTTCTTTTTAATTCTCGGAATCACAATGGTTTTGGCATTTATCCAAGACGTGATAACGG
GTTCTAATTTTCTGCAAATAACAATTAATGTAAAATTTTCGTAAAAATTTATCGGCTTTT
AAAACAAAATTGACTAAAATAGTCGCGAGTTTTTACTGCAATAAAGGAGATTGCAATGAA
TATGAAAACCTTATTAGCACTAGCGGTTAGTGCAGTATGTTCAGTTGGTGTTGCGCAAGC
ACACGAGCATAATACGATACCTAAAGGTGCTTCTATTGAAGTGAAAGTGCAACAACTTGA
TCCAGTAAACGGTAACAAAGATGTGGGTACAGTGACTATTACTGAATCTAACTATGGTCT
TGTGTTTACCCCTGATTTACAAGGATTAAGCGAAGGCTTACATGGTTTCCACATCCATGA
AAACCCAAGCTGTGAGCCAAAAGAAAAAGAAGGTAAATTGACAGCTGGTTTAGGCGCAGG
CGGTCACTGGGATCCTAAAGGTGCAAAACAACATGGTTACCCATGGCAAGATGATGCACA
CTTAGGTGATTTACCTGCATTAACTGTATTGCATGATGGCACAGCAACAAATCCTGTTTT
AGCACCACGTCTTAAACATTTAGATGATGTTCGCGGTCACTCTATTATGATCCACACGGG
TGGTGATAATCACTCCGATCATCCAGCTCCACTTGGCGGTGGCGGCCCACGTATGGCATG
TGGCGTGATTAAATAATTCGATTGTTCGAAACGAAAAGTGCGGTGAATTTTGACCGCACT
TTTTTGCTAGATATTTAGCATTGAGACCTTTGCAATAACATAGGTTACTAAAATTTTATG
CTCAATCTCATTTTCAAAATGCAAAACTTTTCTGATTTTTCCTACTTTTTGCTCAATATT
AGGAAGGTTTTAGGCAATTGAAAATTTTTTGGCGCATTTTTATGCGTCAAATTTCGTTAA
CAGACTATTTTTGCAAAGGTTTCAATTCATAAGTTTCCCGAAATTCCAACATAACCGAAA
CCTGACAATAACCGTAGCAACTGAACCGTCATTCCCGCGAAAGCGGGAATCTAGACCTTA
GAACAACAGCAATATTCAAAGATTATCTGAAAGTCCGAGATTCTAGATTCCCGCTTTCGC
GGGAATGACGAAAAGAGACCTTTGCAAAATTCCTTTTCCCCGACAGCCGAAACCCCAACA
CAGGTTTTCGGCTGTTTTCGCCCCAAATACCGCCTAATTCTACCCAAATATCCCCTTAAT
CCTCCCCGGATACCCGATAATCAGGCATCCGTGCTGCCTTTTAGGCGGCAGCGGGCGCAC
TTAGCCTGTTGGCGGCTTTCAACAGGTTCAAACACATCGCCTTCAGGTGGCTTTGCGCAC
TCACTTTAACCAGTCCGAAATAGGCTGCCCGGGCGTAGCGGAATTTACGGTGCAGCGTAC
CGAAGCTCTGTTCAACCACATAACGGGTCTTCGACAAATATCGGTTGCGTTTGGTTTGCA
CTTCCGTCAGCGGACGGTTGCGGTGGGCTTTGCGCATAATGCCGTCCAGCAACTGATGTT
CTTCCAGATGTTGCCGGTTTTCCGCACTGTCGTAGCCTTTGTCGGCATAGACGGTCGTAC
CTTTGGGCAGTCCTTCCAACAACGGCGACAGGTGTTTGCACTCATGGGCATTGGCGGGGG
TAATGTGCAGTTTCTCGATATAGCCTTCTGCATCGGTACGGGTATGTTGTTTGTAACCGA
GTTTGTAGAGGCCGTTTTTCTTTATCCAACGGGCATCGCTGTCCTTACTCGGTGTGGTTT
GACCGCTGATTTGTCCTTCTTCGTCAACTTCTATGGCCTGACGCTGTTTGCTGCCGGCGG
TCTGAATAATGGTGGCGTCAACGACGGCAGCGGATGCTTTCTCTATTTTTAAACCTTTTT
CGGTCAGTTGGCGGTTAATCAGTTCCAACAGTTCAGACAGGGTATTGTCTTGCGCCAGCC
GGTTGCGGTAGCGGCATAAGGTGCTGTAATCGGGGATGCTCAGTTCGTCAAAACGGCAAA
ACAGGTTGAAATCGATGCGGGTAATGAGGCTGTGTTCGAGTTCGGGATCGGAGAGGCTGT
GCCATTGTCCGAGCAGGACGGCTTTGAACATGGACAGCAGGGGATAGGCAGGACGGCCGC
GGTGGTCTCTAAGGTAACGGGTTTTTTGACGGTTCAGGTATTGTTCGATCAGCTGCCAAT
CAATCACCCGGTCCAACTTCAATAGCGGGAAGCGGTCGATGTGTTTGGCAATCATGGCTT
GGGCGGTTTGCTGGAAGAAGGTGCTCTTGAGAAATCCCCTAAATGTCTTGGTGGGAATTT

AGGGGATTTTGGGGAATTTTGCAAAGGTCTCTAGATGAGTGAAAAAGAAGTGCAGGCTGC
CTAAAAAGACAGAAAAAGTCTTTCCGGCAGCCTGCACTTTGGTTTCATTTCAGTCAGTAA
ACCCAGTAAACGACGGTCTGAAAACGCAGAACGTTACGAAAAAAGCAGCCTACACGCCCA
TCCCCCGCCTTCTACCCGTTCTGTAAATCATACAGATAGCGGTAATATCCGTTCGGCTTC
GCCAGCAATTCCTGCTGTGTTCCCGCTTCCACAATCCTGCCTTTATCCATGGCAATGATC
CGGTGTGCCGTTTTAACAGTGGACAGACGGTGGGCGATAATCAGCACCGTCCGGTTGGCG
CAAATGGCCTGCATGTTCTGCATAATCGCTCGTTCACTTTCATAATCCAGCGCGCTGGTG
GCTTCATCAAAAATCAGAATGCGCGGATTGGTGATTAACGCGCGGGCAATCGCAATACGC
TGCCGCTGTCCGCCCGACAAGCCGGCCCCTTGTTCGCCCACCACGGTGCCGTAGCCTTCC
GGCAGCTCCATAATAAACTCGTGTGCGCCCGCCAGTTTGGCTGCTTCGATAATGCGTTCC
AGCGGCATACCCGTATCCGTCAGCGCGATATTGTCGCGTATGCTGCGGTTGAGCAGCACA
TTCTCCTGCAAGACCACGCCGACCTGCCGCCGCAGCCAGGCAGGAGCGGCCAAAGCCAAA
TCGTTGCCGTCCACCAACACCCGTCCCTGCTCCGGTACATACAGACGCTGCACCAATTTG
GTGAGTGTGGATTTGCCCGACCCCGAACGTCCCACAATCCCCAGCACTTCCCCCGCCCGA
ATCCGCAGGTTCAAATCCTGCAAAATCAGCCTGCCGTCCGCCTTATAGCGGAAATCGACA
TGTTCGAACGTAATCTCCCCCCGGATATCGGGCAAAGCCAAATGCGAAGACGCATTCTCG
GTCGGCGCATTCAGAATATCCCCCAAACGCGCCACCGAAATCCCCACCTGCTGGAAATCC
TGCCACAACTGCGCCAAACGGATAACAGGCGCCGCCACCTGTCCCGAGAGCATATTAAAC
GCAATCAGCTGCCCCACCGTCAGCTTGCTCTCAATTACCAGCCGTGCGCCAATCCACAAC
GTCGCCACCGTCACCAGCTTCTGAATCAGCTGCACCCCCTGCTGGCCGACCACCGCCAAC
TTCGTTACCCGAAATCCCGAAGCCACATAAGCCGCCAACTGATTGTCCCAACGCTGCGTC
ATCTGCGGCTCCACCGCCATCGCCTTTACCGTACCCACCGCAGTGATGCTTTCTACTAAA
AACGACTGGTTGTCTGCATTGCGCGCGAACTTATCGTTCAGACGCGTCCGCAGTATCGGA
CTGATAAATGCCGACCAAAACGCATAGGCAGGCAACGAAGCCAATACCACCCAAGTCAGA
GTGGAGCTGTAATACCACATCACCGCCAGAAAGATAAACGAAAACGCCAAATCCAACACC
GAAGTCAGCGCCTGACCGGTCAAGAAATTGCGAATCTGCTCCAATTCCCGCACCCGAGCC
ACCGTATCACCCACTCGTCTGTGCTCGAAATAGGATAAAGGCAGGGAAAGCAGATGCCGG
AACAAACGCGCGCCCAATTCCACATCAATACGTGAAGTCGTATGTGCAAACAGATACGTC
CGCAAACCGCCCAACACAATCTCAAACAGCGACACCACCAACAAAGCCACCGACACCACA
TCCAAAGTAGAGAATCCCCGATGTACCAGCACCTTGTCCATCACCACTTGGAAAAACAGA
GGCGTAATCAGCGCAAACAGCTGCAACACCACCGACACCACCAATACTTCAAAAAACAAC
CGGCGGTATTTGATTACCGCCGGAATAAACCAGGTAAAGTCAAACTTTGCCAAACTGCCC
AATACCGAAGCGCGGGAAGCAACCAATATCAGTTTGCCCGAATATCTGTTAGAAAATTCG
GCAAAAGACAATACCGCAGACTTATTCGTAACCAAATCCTGTATCAAAAATTGGGCATGC
TCACCCTCACCGTCTGTTTTGGCCAAAATGAAATGGTTGCCGTCATCACACCATACCAAT
GCGGGTAAAGTCGCCATAGCCAAACGTTTAATAGGCTGGCGGACTACCTTTGCCTTCAAT
CCCAAAGATTTGGCGGCTAACAGCCATTGCGTTTCATTTAAATCGCTCTGTGCGGAAGTA
CAAAATTCATGCTGTATATCGGCAGGATTGGCGGCAATGCCGTGGTAATGGGCGAGGATG
ATGAGGGCGGAAAGGGCGGGGAGCGGTGCGGATACGATAGACATAAATAAAATATAGTTA
GATTGGATGTGGATAACGGCTGGCTGGAAAAGGAATATATTAAGTAGAAAGAAATATATA
AATAAAACAGCAGAACGCATTGTAAGGATATATATGGGAATTGTAAAGAGAAAGTATGGA
AAAGTTCTCGTTTCAGGAAGGTAAAACGGCTTAGGAATCGAGTTAGATGAGGATGCCTCG
CACCTCTCGTGCCTCCTGCATACCGTTAAGGCACAGGGTTAAGGTGCAGGCTGCTCCGAA
CTCTGTTGCGGTCGGGTAATGTTATTTTTTGTGTTTCAGGCAGCCTGAAATATCTGTATA
TTTTTGTTTTAAATAGATTTTAAAGATTGATAACTGTTCTTGACGATTTTTCAAGAAAGG
AGTAAATTTCAAGAAAGGAGTAAAGTGACTTATTATCAATGACAAGCAACGCGCGAAGTG
ACAAGGAAAACTATCTACTTAAATTCTAAGGAGGCTTCGAATATCATAAACCAATCAGAA
ACATAGAGATAAAAATTATGTACAAATATAATCCTCTTATACAATTTATTGCACAGTTGA
TTATGTCTTATGGAGCAAGCGTAGGGTGGGCACTTGCTGCCCCACGCGTTTCATATTTCA
AGGCAGCCTGAAACCGTGTGGGCATAAATGCCTACCCTACATCCCAAAAAACAAGCGCAG
CCTGCGTGTGTAGGGTGCGAACTTTCGGCAGGTAGACACGCAGTTTTATATTTTCAAGCT
GAGGGATGCTTAAGAAAAGTACAAAACATTAAAAAATAAGGGGCTGTACTAGATTAGCCC
TAAATCCACACCAATCCCGCAAGATTTTTAGCTGTCGGGACGGTGTGCCGAAGTTAAATC
GAAATTCGCATTCTTTCAAGAACAGCGGGAAAGATTTGCGATCAATTCCGTTCTATTTGC
GCAAGACGCGTTTTGCCTGATTCCAAAAGTTCTCAATGCCGTTTATGTGGTTCTGACGGT
CAGCAAATTCCTTGGAATGGTTGATGCGGTAATGGATAAAACCGCTTACGTCCAACTTGT
CGTAACTGCTCAGGCTGTCGGTATAAACAATGCTGTCCGGCATGATTTTCTGTTTGATAA

CAGGCATTAAAGTATCGGACTTGGCATTATCTACGACAACGGTATAGACCCGTCCGTTAC
GTTTCAGAATGCCGAAGACAACCACTTTTCCTGCCGCACCGCGACCACGTTTGCCTTTAC
GCCGTCCGCCGAAATAGCTTTCGTCCAACTCGACAGAGCCCTCGAAAACCTCATTGGCAG
CCAAGGCCAGATAATGGCTGATGACCATACGGATTTTGCGGTAGAACAGGACTGCCGAAT
TGGGATGGATACCCAAAATATCGGCAGCAGAACGGGCGGTAACTTCGAGTACAAAAAAAC
GGAGCAGCTCTTTCTGTACTTTTTTCTTTAATTTGCAGTGTGTTATCTTCATATTTCGGG
GGTAACATATCTGCTAATCTAGTACAGCCTCAAACAAAAAAGAGAAATTTTAATTTCGCT
AAATCGCATAAAGATTAATCAAGAGTATCATTAAATGATATGAGTGAGCATCTATAATGC
CAAGAAGAGTTGTTAAGACATAACGATTATTGAAATAGATTGTAAAATAGATACTTAGAT
AGTCTGAAAAACGGATTTGTGAAACTTTTTATTACGCGCCATCATTTGAAAATGAAACTT
AAAAAACACTTATCATAATAAATATTTTCTTTACGTTGTTTGCTAATAAACTCAGTGCAA
TATCAGCGCAATATTTTATGGAAATTTTATGGATAACAAAAAAGAATTTATTAATAATTT
AACAAATAGGTATATGTGGATCTATCCATTGGTCTTAAATATTCTATTTCTACCTTTTTA
CCAGTCCTACCAATCTTTTTTTATTGCGCTTGGTTGTTTGTTTGCACTGGTTAGAAAAAT
GCAAAGCTTAGATTTTAAATTACAAAATCATATTGTATTGTTAAATATAAAAAGTGCTTG
GGCAGATAAAAAAGTATTTTTGATTAGGATAGTAGTGTCATGGTTGGCAGTAATGGAAAT
ATGGATGTGTTTTATTTCGGAATCATCAACGTGGGTATGCGGTGCTTTTTGTTTAAATAG
TGAAATATTGGAAAAAATTTTTCGTGGCTTTGGTTATTCTGGTAGTTTATATTTTTTATT
TATATTGATGATTGATCTCAATAAGTTAAGAGAGAGTATTTGAATTTCATTTTTTGTTTG
ACTTAAACTCAAGGAGAGTAACAATGATTGGTAGTGGTGATACTAAACAATGCAAAAAAT
TTTCTGCGTGTGATGGAAAATACCACGTCTACGATCCCCTCGCCCTAGACTTGGACGGCG
ACGGCATAGAAACAGTCACCGCCAAAGGCTTTTCAGGCAGCCTGAAGACTGAGAGAGTGA
ATACGATGAGTATACACTCTATGCCACTAAATTGATATTCACTAAATCATACCAGCTATA
TTTTATTTAATGAGACATATGAAAAATAAAAATTATTTACTAGTATTTATAGTTTTACAT
ATAGCCTTGATAGTAATTAATATAGTGTTTGGTTATTTTGTTTTTCTATTTGATTTTTTT
GCGTTTTTGTTTTTTGCAAACGTCTTTCTTGCTGTAAATTTATTATTTTTAGAAAAAAAC
ATAAAAAACAAATTATTGTTTTTATTGCCGATTTCTATTATTATATGGATGGTAATTCAT
ATTAGTATGATAAATATAAAATTTTATAAATTTGAGCATCAAATAAAGGAACAAAATATA
TCCTCGATTACTGGGGTGATAAAACCACATGATAGTTATAATTATGTTTATGACTCAAAT
GGATATGCTAAATTAAAAGATAATCATAGATATGGTAGGGTAATTAGAGAAACACCTTAT
ATTGATGTAGTTGCATCTGATGTTAAAAATAAATCCATAAGATTAAGCTTGGTTTGTGGT
ATTCATTCATATGCTCCATGTGCCAATTTTATAAAATTTGCAAAAAAACCTGTTAAAATT
TATTTTTATAATCAACCTCAAGGAGATTTTATAGATAATGTAATATTTGAAATTAATGAT
GGAAACAAAAGTTTGTACTTGTTAGATAAGTATAAAACATTTTTTCTTATTGAAAACAGT
GTTTGTATCGTATTAATTATTTTATATTTAAAATTTAATTTGCTTTTATATAGGACTTAC
TTCAATGAGTTGGAATAGTTTTGGTAATTTTATGAGCGCACGCTCATCCGCGTTAGCAGA
ATTTGGAAATATGGTTGCTAATTTAGTTTCTGCAAAAAATGAGAAAGATATCTCGAAACG
TAATGAATATTACAAACAAGCTGGTTATAGTGCATTATTAGCATTTGGTAATTTGGCTAG
TAATATTGCACCAGGTAGTACGTCATCGCATATTGTAAACGGAACAAATGCCTCTGTGAT
TGCAAGCCGTCTCTCTGGAAATATATCTTCAGCTATTCAGGAGCATAAAGATGGTAAAGT
TAATATCAACCGTTTTCAAAATATTTTAGCGGATTTATATTCATTGGGAGGGTTAGGAAG
TACATTAATAGAGAAGAATGGAAATATGCAGAGTTGGGGGATTCCATTAGCAATTGCTGG
AGATATAATTGCAGCAACGGCTATTGCCACAGGAGATACTGGTACGATATCTACAGAGGA
ATTTTATAATTTTGACAACTGGAAAGGTTTTGGGTATGAGCTATTTGAAGACTGGTCTCG
TTGGGTATACGACTGGCTGCCCGACGGCTGGAATCTGTGGAAAGAATTGGACAGAAACCG
TTCAGGCCAATACCACATCTACGACCCCCTCGCCCTAGACCTAGACGGCGACGGCATAGA
AACAGTCGCCGCCAAAGGCTTTTCAGGCAGCCTCTTCGACCATAACGGCAACGGCATCCG
CACCGCCACTGGCTGGGTTTCTGCCGATGACGGTTTACTCGTCCGCGATTTGAACGGCAA
CGGCATCATCGACAACGGCGCGGAACTCTTCGGCGACAACACCAAACTGGCAGACGGTTC
TTTTGCCAAACACGGCTATGCAGCTTTGGCCGAATTGGATTCAAACGGCGACAACATCAT
CAACGCGGCAGACGCCGCATTCCAATCCCTGCGTGTATGGCAGGATCTCAACCAGGACGG
CATTTCCCAAGCTAATGAATTGCGTACCCTTGAAGAATTGGGTATCCAATCTTTGGATCT
CGCCTATAAAGATGTAAATAAAAATCTCGGTAACGGTAACACTTTGGCTCAGCAAGGCAG
CTATACCAAAACAGACGGTACAACCGCAAAAATGGGGGATTTACTTTTAGCAGCCGACAA
TCTGCACAGCCGCTTCACGAACAAAATGCTATCCATTAGCCATGTTCGGGAAAACACGAT
TTCCCCGTTTGTTTTAGGCTGTCTAAACAAATAACCATAAATGTATATCATTATTTAAAA
TAAATAAAAGTATTTAACTATTATTGACGAAATTTTAGAGAAAGAGTAGACTGTCGATTA

AATGACAAACAATAGTGAGAAAGGAAATATTTACTATCCGAGCACAGAGCATATTTTAGG
TAGCCTGTAACTGTTCCTGCTGGCGGAAGAGGATGAAGGTTGACTTACCCGAGAATAAAT
GTCCTGTTGTGTGATATGGATGCCATGCCGCGAAGCAATTGATGCAATCACGGCAGTCCT
ACTTGAATGAAACCTGTCGTTGCAGAATTTGAAAACGCTATTTTTAAGAAAGGATAAAGG
GAGAAAGAATTTTTGGTTTTTAAGCTGCATGAAACCGTGTTGGAATAAATGCACACCTAC
GATAATTAATAATTTTCGTTTTTATTCTACAAGCTATTTATATATGATTGCTAAAAGTT
TATTTTTTAGATGCCAAAAAATATATTTTATATACTTCATATTGTTTATATGTCTTTATT
TGAATATATCTTACGATGGGGAAATATTTATATATTTTATAATAAATTTTACTCATTTGC
TAATATGTCATGGAATATTACTTGTATTTTGTAGAATTTTTCCATATGAAAATATTCCAT
TTACTATTTTTCTGAACTTTATTAGTTTATTTTTAATATTTTTACCTCTTATATTTACCA
TAAGAGAGCTAATTGATTCATATTATATTGAGTCGATAATTAATTTATTCTTAATTTTAA
TTCCTCACGTTATTTTTTTAATTTACTTGAAAGGAAAGCAGATATGACATCTGCAAATTT
TAATATTAACGGTTTTGGAGATGTGAAATTAACACCCTATTCACCACTCTTGGGATATAA
AGCTTGGGATTCATTTATTGGTTCTATTCAATCCTTATCTGATTTAATCTATAATGTGGA
TAACAATAGAAATAAAATGGAAATTACTGTTAATAATGCTATTCAAGCTGCAGATAGCTT
TTTAAGCAGTAATTGGAAGAGATAACAAAATAACAAAATAACAAAATAACAAATACTGCT
TCTTTACTTGCATCCTTCGATAACATTTTTTAAATTTAAGAAATGTATCTCGAGATATAC
GAGAAACAGGAAAATTTAAACCTAATGATATTCAACAAGCAATTGGTGATATATTCATTG
CTGCTGGTGATGGATTACAATATATAAAACAACAAACAGAGGCGATGGCTCAAAGCAAAT
TCTTACCAACTAAATTAAAAACTGGTTTAAATGATGTCCTTAATTCTAGAATGCTAAAAT
CCTCTACTGTTTTACAGCATGAATTGAATTAAATAAGGATTATGGAAACGAGAGGCTTGG
CGAATCTATAATGAATATAGATGATTTTACACCAAGTAAGATAGCAAACTTTTTTGCGGA
TCCTGATACATACAGCAATGTATTAGAAGAAGTATCTAGGTTTATATATTCCTTAGTTCC
TGATGATGCAAACCCTTGGAAAGGGGGCGAAGATTATATTGGACGAGGGATAAGTGAATG
GGGAGAGTTACTGGAAAAATGGTATAAACAAGATTTTCTCCCTTATCTTGAAAAGAATGG
GACCAATTTCCGAAATTTGAAGATTGGCTGCCTGAATTCCCTGAATGGGCAAGAGAGTGG
TTGAAATTAGCTCTCAAACGTTCAGGCAAATATAACGTTTACGATCCCCTCGCCCTAGAT
TTGGACGGCGACGGTATAGAAACCGTTGCCACCAAAGGCTTTTCAGGCAGCTTATTTGAT
CACACCAACAACGGCATCCGCACCGCCACGGGCTGGATTGCTGCATATGACGGTTTTCCT
GTGCGCAAATTAAACAGTAACGGGGGCATTATTAGCACGACAGATACCATATTCCAATCT
TTGCATACATGGCTTGATCATCAACCAAGATGATATTTCCCAAGCACAGCATGATGCATG
CCATTGAAAAATATAGAAAATTAATTGAAAGCTTAAATGGATATTGAAATGAATGATCAC
ATAGTACAAATTATAAGAAGGTTCGGGCTAGGTAGGATATTTTTTTATTCGTATAGCAAA
TCATCTATAATAATTTTTTCTTCGTATGTTGTTTATTATATAATTTACAATTATCAATTT
AATTACCTTTCGCTTTTAATTTATTTATTACCAATATTGTGCAGTATATATATGTTTATA
TTTTTTTTAGGGAAAACTAAGGATACATTAACGACAGAGCGAAGAAAAAAATTTTTTAAT
TCTATTTTTCCACTTAGAATTCTAATGATAATAGGTTCTGAGAAAAAGAGGTTAGGCATC
GGTAGTTTTTATTTGCTAAACCTACTATGGATTATTTGGTGTCTTATGATTCATAGAGAA
CAAGTCCCATTAAATAACTTAACCCTCCTATTATCCTTCATATTTTCATTATTTTTTTTA
TTATGTGATTTTGTTTTATTATTGAATGTTTATGTTTATTTTTTTAAATTAAGAGAGGCT
TAATATGGTTAATCAAATCAAATCTGATAATAATTCAGTTTCTATTGAATTTATATAAGA
TTTTATAACTGCAAGTACGGATGTAATTAATCTGAGTTACGAAAATTTTCGTAAAAATTT
TTATACACAAATGTCAACTGATTCTACCAATTATGCAGCCAAACATGAAAGTTTAGGAAA
ATCGGTACAACGTGAATTACAAAAAACACAAAGTCAGTTGAGACAAGTTGTAAGAAAAAT
GCAGAGTAAATATAATATAAATAATAAAGCACGAGTAGCAGAAATATCTTTGTTAAGGCA
AATGCAAAGCCAATTTTCTCGAAAATATGTAAACAAAAATCTTGGTAACAGCAACACTTT
GGCTCAACAAGGCAGCTACACCAAAAAAGACGGCACAACCGCGCAAGCAGGCGATTTGCT
GTTGGCTGCTGACAACCTGCACAGCCGCCTCACGGACAAAATGCTATCCATTAGCCATGT
TCGGGAAAACACGATTTCCCCGTTTGTTTTAGGCTGTCTAAAACAAATAACCATAAATGC
ATATCATTATTTAAAATAAATAAAAGTATTTAACTATTTTTGACAAAATTTTAGAAATAG
AGCTAGAGTTTTAGTTAAGTAGAAATTGATAGTGCTTCAAGGGAAGTATTCTCTATGTTT
GCATTAAAGGGGGTCTGATAAAGCTATTATTCATTACTATGGACTTTTATTTCATTATTT
TCAGGCGGAAATCTCATAGCCGTTTTGAATTTTTCTCTTCCTTATTAATTATACAAATAA
TTAGTATATTCTGATATGGATTTTTTGGAAATTTTTATTATGTCTGCATTTAGAAAAATA
TTATTAATAATATCTTGCCTATTGATTGCTAGCTGCAGTTTTGTTGAAACTATTTTTTAT
ATGGCTATTAGCCCAGAACCTGTTGTGGTAGACTTTCCTCTTGGTAAAAAAACAAAAAGA
TCTATTGAACTCAAACAGAAAATTGGTAAACCTTATGCAATATCGTTAGGAACTAATTTT

ATACATTATGATCCAAAACAGGGGGAGAGGTGGATTGATGATAAGTTAAACTATCCATAT
AATATATCGGTTAAAATATTTAAAGTGGAAGAAGATGGTAAAAAACTTATTATAGATGAG
TTGCTTACAGAGAGAAGTAGAAAATTAGGAGGCGGAGTATTTGGAGCTGGGGAAAATAC
AGTATGCATATTTATGATTTTTATTTGCCGGAAGGGGAATATTTATTTGAGATTTCTGAT
AATAGTGAATATATTCCACTTTACGATGAAATAAATAATTCTATAAGAATAGTAGTTAAT
GCACGAATTCAGTAAATTTTTCTAGAAATGTGGGGTTACTTATGGCTGATTATTATGCGA
TAACTGTAAAATTTGCGAAGCAGGGTACGCCACTGAAACAAGAGGGGGTGTATCCAAGAC
GGGTACGTTTGGGTTGAACTGTATTCGGCTAGAGATAAAAAAATCGGGGCTGTACTAGAT
TAGCCCTAAATTCCACACCAATCCCGCAGGATTTTAAGCTGTTGAGACGGTGTGCCGAAG
TTAAATCGAAATTCGCATTCTTTCAAGAACAGCGGGAAAGATTTACGATCGATTCCGTTG
TATTTTCGCAAGACGCGTTTTGCCTGATTCCAAAAATTCTCAATGCCGTTAATGTGGTTC
TGACGGTCTGCAAATTCCTTGGAATGGTTGATGCGGTAATGGATAAAACCGCTCACGTCC
AACTTGTCGCAGCTGCTCAGACTATCGGTATAAACAATACTGTCCGGCATGATTTTCTTT
TTGATGACAGGGAGTAACGTTTCAGACTTGGCATTATCTACGACAACGGTATAGCCCCGT
CCGTTGCGTTTCAGAATGCCGAAGACAACCACTTTTCCTGCCGCACCGCGACCACGTCTG
CCTTTACGCCGTCCGCCGAAATCGCTTTCGTCCGGCTCGACAGGGCCCTCAAAAACCTCA
TCGGCAGCCAAGGCCAAATGATGGTTGATAACCGTGCGGATTTTACGGTAGAACAGTACT
GCCGAATTGGGATGGATACCCAAAATATCGGCGGCAGAACGGGCGGTAACTTCCAGCACA
AAAAAACGGAGCAGTTCTTTCTGTACTTTTTTCTTTAATTTGCAGTGCGTTATCTTCATA
TTTCGAGGGTAACATATCTGCTAATCTAGTACAGCCCCAAAAATATACCAAAAACAGCAA
AACAAATTGTAAGGATAGGTATAGGCTTTGTAAAGGTAAATTGTGAAAAAAGCAGTTTTT
TAAACGAATGAAACGGCTTCGGGCTGAAATATATGCTGATGCCCTGTCCTTCCCGTATAT
CTTGTGTGTTGTCAAAGTGCAGGCTGCTTTGAAATCGGTATTGCCATCTATGAACCACCA
CTTTGTTTTATTTCAGCGGGCTTGAGATGTGTATAAGAATATTGTTTTGAATAAATTTAA
AAAAATGATAATCGTTATTGACGATTTTTAAAGGAAAGCGTAGAGTGCCAATTCTATGAA
GCAATACGGTAAGTAACAATGAAAATATCTACTGCTTGGGTATAGAGCATATTTCACAAC
CCGTAACTATTCTTGCGGAAACAGAGAAAAAAGTTTCTCTTCTATCTTGGATAAATATAT
TTACCCTCAGTTTAGTTAAGTATTGGAATTTATACCTAAGTAGTAAAAGTTAGTAAATTA
TTTTTAACTAAAGAGTTAGTATCTACCATAATATATTCTTTAACTAATTTCTAGGCTTGA
AATTATGAGACCATATGCTACTACTATTTATCAACTTTTTATTTTGTTTATTGGGAGTGT
TTTTACTATGACCTCATGTGAACCTGTGAATGAAAAGACAGATCAAAAAGCAGTAAGTGC
GCAACAGGCTAAAGAACAAACCAGTTTCAACAATCCCGAGCCAATGACAGGATTTGAACA
TACGGTTACATTTGATTTTCAGGGCACCAAAATGGTTATCCCCTATGGCTATCTTGCACG
GTATACGCAAGACAATGCCACAAAATGGCTTTCCGACACGCCAGGGCAGGATGCTTACTC
CATTAATTTGATAGAGATTAGCGTCTATTACAAAAAAACCGACCAAGGCTGGGTGCTCGA
ACCATACAACCAGCAAAACAAAGCGCACTTTATCCAATTTCTACGCGACGGTTTGGATAG
CGTGGACGATATTGTTATCCGAAAAGATGCGTGTAGTTTAAGCACGACTATGGGAGAAAG
ATTGCTTACTTACGGGGTTAAAAAAATGCCATCTGCCTATCCTGAATACGAGGCTTATGA
AGATAAAAGACATATTCCTGAAAATCCATATTTTCATGAATTTTACTATATTAAAAAAGG
AGAAAATCCGGCGATTATTACTCATCGGAACTATCATAGGTATGGAGAGAACGATTACAG
CACTAGCGTAGGTTCCTGTATTAACGGTTTCACGGTACGGTATTACCCGTTTATTCGGGA
AAAGCAGCAGCTCACACAGCAGGAGTTGGTAGGTTATCACCAACAAGTAGAGCAATTGGT
ACAGAGTTTTGTAAACAATCCAAGTAAAAAATAATGGGGCTGTCCTAGATAACTAGGATA
AACTCGATTTTACTAATTGTTTTAAAATGGAACAAGAACTTTTATCTCACTGTTGTTAAA
ACGCCATTCGCACTCCTTTAAATACAGCTCAAAATGCGCTTTGGGAATGCCGTTAAACTT
GCGTAAATGACGTTTTGCCTGGTTCCAAAAGTTCTCAATTCCATTAATATGGTTTTGTCG
TTCAGCAAAATGTGTGCTGTGATTGATACGAAAACGAAGTTTCAGCGAAGCTAAAATGGC
TAAATTCGCGCACATCTAATACATCATAGCTACGATAACAATCCGTATAAATAATGCTGT
CAGGTTTCACTTGTTCACGGATAATAGGAAATAAAGTAGCGGTTTGAGTATTCGGTACTG
TAACCGTATAAACCTTACCATTTCGCTTCAAAAGACCGAATACGGCGACTTTACCGGCAG
CACCGCGACCGCGTTTGCCTTTGCGTTGTCCGCCAAAATAACTTTCATCTGCTTCTACTT
CGCCATCAAACATTTCCAAATGCGGACTGTTTTGATAAATAAGTAATCGTAAACGATGAA
AATAATAGGCTGCGGTATTTTTATTAACGCCTACTAACTCTGCTGCCGTTCTTGCAGTTA
CACCTGTGACAAATAGCTCAATGAGTTTATTTTGTTTATACTGGCTTAGACGACTTTTTC
TCATAGGGATAATTCTAACTTAATTTGAATTTCCCTAGTTATCTAGGACAGCCCCTATTC
TTTAACTAATTTCTAAGCTTGAAATTATGAGACCATATGCTACTACCATTTATCAACTTT
TTATTTTGTTTATTGGGAGTGTTTTTACTATGACCTCATGTGAACCTGTTAATGAACAAA

CCAGTTTCAACAATCCCGAGCCAATGACAGGATTTGAACATACGGTTACATTTGATTTTC
AGGGCACCAAAATGGTTATCCCCTATGGCTATCTTGCACGGTATACGCAAGACAATGCCA
CAAAATGGCTTTCCGACACGCCAGGGCAGGATGCTTACTCCATTAATTTGATAGAGATTA
GCGTCTATTACAAAAAAACCGACCAAGGCTGGGTGCTCGAACCATACAACCAGCAGAACA
AAGCACACTTTATTCAATTTCTACGCGATGGTTTGGATAGCGTGGACGATATTGTTATCC
GAAAAGATGCGTGTAGTTTAAGCACGACTATGGGAGAAAGATTGCTTACTTACGGGGTTA
AAAAAATGCCATCTGCCTATCCTGAATACGAGGCTTATGAAGATAAAAGACATATTCCTG
AAAATCCATATTTTCATGAATTTTACTATATTAAAAAAGGAGAAAATCCGGCGATTATTA
CTCATTGGAATAATCGAGTAAACCAGGCTGAAGAAGATAATTATAGCACTAGCGTAGGTT
CCTGTATTAACGGTTTCACGGTACAGTATTACCCGTTTATTCGGGAAAAGCAGCAGCTCA
CACAGCAGGAGTTGGTAGGTTATCACCAACAAGTAGAGCAATTGGTACAGAGTTTTGTAA
ACAATTCAAGTAAAAAATAATTTAAAGGATCTTATTATGAATGAGGGTGAAGTTGTTTTA
ACACCAGAACAAATCCAAACCTTGCGTGGTTATGCTTCCCGTgGCGATACCTATGGCGGT
TGGTATCCGAGCT

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 4>:

gnm_4

CGGGGCTTACAACATCGGCGCAATCGACAAAAAGACCATGCGCGACTTTGACAAGTCCTG
CCTGACCGAAATCAAACCGTTGAGCGGCGGAGACATCAAGGCAATCAGGGAGAAGGAGGC
ACTATCGCAAGCCGCTTTCGCCATCTATCTCAACGTGGGAAAAAATCACGTTTCGGCTTG
GGAGCGGGGCGTTAAAAAGCCGAGCGGCGCGGCGTTGAAGCTGCTGACCATCGTCAAAAA
CAAGGGCATCGAAGCCATTGCGTAGCCGACTTGGCAAACGGCAAAATCAGCAAGTTCACA
ATAGACGCGCTGCTGAATATGCCTGCCAAGACAGGCAAGACCGCCGAACTGAATATCAGG
GCGTAGCCGCATAAATGCCCGACCGCATCAAACCAAGCCGAAACGGCGGCGGTGCAGACG
ACATAGCCCGACAGCAAGGCACGGCGCAGACGGGCGCGAAACCCGAAACATCACCGACCG
CGAGGTACGGGGATTTTTTGCGCCCGTTGCAGGGGGGGATTGGATTTAAGCGGCGCGGGC
TTGAAGGCAAAACGGGTGGGGCACAGAACTGTTTAAATGCAGTCTGAATCTCAAACGATT
TCAGACGGCATTTTGAAACAATGGCTCAAATTCTCGATCCCCTTCCCTTAACGCCGACGT
TTTTTATTAACGCGCCCCTTATTTCTGACACTTTGCTCATAAACCGGCATAACGGTCGGC
AACAACCGTTTTAGATTTTCTATACGGGCATTGTTTGTCGGATGAGTAGAGGTAATAGCA
TAAATAAAGCCGTTTTGGTCGTTTTCCTGATTCATTTTTTCCCAAACCCTGACAGCGGCC
GCCGGATGATAGCCTGCCTGCGCCATCAACATCATTCCCCCCTCATCGGCTTCTTCTTCC
AAGCTGCGGCTATAAGGCAAGGTAAGACCGTACGTCCCCAAAATATCCATACCCAATCCG
ACCAATTCCGGATTAGTATCCGGTTTTTTGTCTAATATAATCTGCGTGCCTATCTGCGCC
GCCGTATTGGTCAAGATTTGCTGCCCGACCTTATTTTTACCGTGTTCATGCAGGGCGTGC
GTCATTTCATGCCCCATAATGGCGGCAATTTCGTCATCGGTCAGCTTGAGTTTGTCGACT
ATCCCCGTATAAAACGCCATTTTTCCACCGGGCATTGCCCACGCGTTCAGCTCATCGTTT
TTGAAAACCGTCATTTTCCAGTCAAACTTATGGCTGGTATTATTTGCCGCATCGGCATAA
GGCAGCATACGTCGAAATACTGCCTGCACCCTGCGGGCTGTTCTGGATGTGGTATCGACA
TTGCCGGCAGACTTGTTTAACTCAACCGTTTTCATATAATCTTTGGCAGCCGCAGCGTTC
ATTGTGGCGGAATCATGACCGTAAACATCAGCAACGACCGCACAAGCCCCCAATACCGAG
ATTACTGCCGACAGGCAGAGTATCCGTTTAAAGGAAGGAAGGAAGGAAATTTCATATTTA
GGTTTACTCCTTAAAAAATTAAATTTCAAAAAAATGCCGTCTGAATCCAAAACGGATTTC
GGACGGCATCTTAACATTGTTTAATGTTTTTAAAAAGATTTACACCACGATGTTCTCCAG
TCTGCCCGGTACGGCGATGATTTTCTTGGCAGGCTTGCCTTCTATGAATTTCACCGCGCC
TTCAGCGGCGTATTCGGCAGCTTCTTCAGCCGGTTTGTCGAAATCACGCATAAATTGCCA
ATAATTCTCCAACTTTTTTACGGCTGCTGCTGCCTTTTGCGGCAATATTGCGCTGAACTT
CAACTGTTTTCAAAATGGCAGAAGAATAAATATCCCTTGTGAATTCAGTATCATGATTTG
AAATCAAAATACCTTGGGAGTTGGGCGCAATTTATTGATTTTTTGTAAAGTCCGCGACCA
ATGAATTCGATCGTATTTTGGTCGCGCAGAATTTGCAACTGTTGGCGGATTTTGTCTCTG
ATATGGTTGTTTTGGGGAAATTGGATGGATAGTTTGTTTTCAAATTCATACATTTGCGAC

AATGTGAATTCTTCGGGGAGTTGGTCGATACATTTCATAACAGCCAGAAGCCAGCCTTTG
CGCTCCGCATTTTGGTTGCGTAAAAACAAATTGGATTGCCATTTTTTCAGAACGGTTTCG
GGTTCGATAATGCGGGAATTGTCTATTAAGAATATTTTGCCGCTTTCAGGCAAAGGGGCG
AGATTGATAGAACACATAATGTGGTTCGGCCGGTTTTTAATGCCTTTATTTCTGGGAATA
ATCATATCCGGCGTGATGAAATGTTTGGGTACAAGCACCAATTGCCGTATGGAGTAATCC
GCTTTTTTATATGCAAGAAAGAAAAAGTTGGGGTTGGTATCTGACCGGATGCGCTCCAAC
ATGGTGTGATATGCACCGTCAGGCACGCTGTTGCCTATGGTTTTTTGATTTTTACTCTTT
AATTCATATTGCTCGTGGCAATTTGGGCAAAAGAGGTCTGCAACAGGTTTGTTATTGGCA
AATCTCTGCATCGGCTTGCTTCCGCAACAGGGGCAGTAGCCGTTTTTTTCCAACCAAGCC
TCGCTCATTACACGGATTTTATGGGTTGCTTTATTTTGTTGCTTTCCCAATTCGGTATCG
AAAAATAAATTCATGTTTTGGATTTTGAGATTTCAGTTATTCGGGGTTCGTCATGCAGAC
AACACAATCCACCTTAAAAAGGCCGTCTGAAACCCTGTTTCCAAGTTTCAGACGGCCTTT
ATCCGTGTGGCTAAACCTTAAAAGCGGTTAGACGACGATGTTCACCAGTCTGCCCGGTAC
GACGATGATTTTCTTCGCCGGTTTGCCTTCCATGAATTTCACCGCGCCTTCGGTGGCGAG
TGCGGCGGCTTCGAGGTCGGCTTTGGATGCGTCGGCGGCAACAGTGATTTTGCCGCGCAG
TTTGCCGTTGACTTGAACCATCACTTCGATTTCGGATTTGACCAAGGCGGCTTCGTCGAC
TGTCGGCCAGCCTGCTTCCCACAGTTTCGCGCCGTTCAATTCGCTCCACAGGGTTTCGCA
GATGTGCGGCACGATGGGCCACAACAGGCGTACGGCGGTTTCCAATACTTCTTGGGCGAC
GGCGCGTCCTTGTTCGCCGCCGGTGTCGGTTTTGTCGTATTGGTTGAGCAATTCCATCAC
GGCGGCGATGCGGTGTTGAACTGCTGGCGGCGGCCGTAGTCGTCGCTGACTTTGGCAGT
GGTCGCGTGCAGTTTGTGGCGCAGGTCTTTGAGTTCTTTAGACAAACCGTCTTGGCTGCC
TGCGAACGCTTTGACCGCTTCGCCTTGCTTCAAGTATTCGTAAACGGTACGCCACAGGCG
GCGCAGGAAGCGGTGTGCGCCTTCGACGCCGCTGTCGCTCCATTCGAGGGACTGTTCGGG
CGGTGCGGCGAACATCATAAACAGGCGGGCGGTGTCCGCGCCGTAGGCGTTAATCAGTTC
TTGCGGATCGACGCCGTTGTTTTTGGACTTGGACATTTTTTCCGTGCCGCTGATGACGAC
GGGCAGCCCGTCGGCTTTGAGGACGGCGGAAATGGGGCGGCCTTTGTCGTCGAACGTCAG
CTCGACATCGGCGGGGTTGATCCAATCTTTGCCGCCTTTGTCGTTTTCGCGGTAGTAGGT
TTCGCAAACGACCATGCCTTGCGTCAGCAGGCGTTCAAACGGTTCGTCAACATTGACTAG
ACCTTCGTCGCGCATCAGTTTGGTGAAGAAACGCGCGTACAAGAGGTGCAAAATCGCGTG
TTCGATGCCGCCGATGTATTGGTCGACCGCGCCCCAGTATTTCGCGGCGGCAGGATCGAC
CATGCCGTCTGAAAATTTTGGCGACATGTAGCGGAAGAAATACCAGCTCGATTCCATGAA
GGTGTCCATGGTGTCGGTTTCGCGTTTCGCCGCGCCGCCGCAGCATGGGCAGGCAGTTTC
GTAAAACTCGGGCATTTTTGCCAGCGGCGAACCCATGCCGTCGGGTACGACGTTTTCAGG
CAAAACGACCGGCAATTGGTCGGCAGGGACGGGTACGTCGCCGCATTGTTCGCAATGGAC
GATGGGAATCGGGCAGCCCCAGTAGCGTTGGCGCGAAATGCCCCAGTCGCGCAGGCGGTA
TTGGGTTTTCGGTTCGCCCGCGCCTTGGCTTTGCAGCTTGGCGGCGACGGCGTCGAATGC
CGTCTGAAAATCCAAGCCGTCCAAGTCGCCGCTGTTGACCAATACGCCGTTTTCTTTGTC
GCCGTACCATTCTTGCCATTGGTTTTCGTCAAATGCGTTGTCGCCGACGGCAATGACTTG
TTTTTTCGGCAGATTGTATTTGGTGGCGAACTCAAAATCGCGTTCGTCGTGCGCCGGAAC
CGCCATCACCGCGCCGTCGCCGTAGCCCCACAATACATAGTTGGCAATCCACACTTCCAG
CTTGTCGCCGTTGAGCGGGTTGACGACGTAGCGGCCGGTCGGCACGCCTTTTTTCTCCAT
CGTCGCCATATCGGCTTCGGCAACCGAACCGGCTTTGCATTCGGCAATAAATGCCTGCAA
TTCGGGTTTGTCGGCGGCTGCGGCGGCTGCCAGCGGATGCTCGGCGGCAACGGCAACATA
AGTCGCACCCATCAGCGTGTCGGGGCGGGTGGTATAAACTTGCAGGAATTTCGCGTAATC
GCCTTCCAAGCCTTGTTTGCTGTCGTCTGAAACGGCGAAGCGCACGGTCATACCGCGCGA
TTTGCCGATCCAGTTGCGCTGCATGGTTTTGACTTGTTCCGGCCAGTGTTCCAGCTTGTC
CAAGTCGTTGAGCAGCTCTTCGGCGTAATCCGTGATTTTGAAGTAATACATCGGGATTTC
GCGTTTTTCGATCAATGCGCCGGAACGCCAGCCGCGTCCGTCGATGACTTGCTCGTTGGC
AAGGACGGTTTGGTCGACAGGGTCCCAGTTTACCGTGCCGTTTTTGCGATAAACGATGCC
TTTTTCAAACAGCTTGGTAAACAGCCATTGTTCCCAGCGGTAGTATTCGGGTTTGCAGGT
TGCGGTTTCGCGCGCCCAGTCAATCGCAAAACCTAGGCTTTTGAGCTGGGTTTTCATGTA
TTCGATGTTATCGTACGTCCAAGCGGCAGGGGCGACGTTGTTTTTCATCGCCGCGTTTTC
CGCCGGCATGCCGAACGCGTCCCAACCCATAGGCTGCATGACGTTGAAGCCGTTTAAAAG
TTTGAAGCGGCTCAATACATCGCCGATGGTGTAGTTGCGCACATGCCCCATGTGCAGCTT
GCCGCTGGGATAGGGGAACATGGAGAGGCAATAATATTTGGGTTTGGAAGCGTCTTCGGA
GACGTTGAAAATACGGGCGTCGTCCCATTTTTTCTGCGCCGCAGGCTCAATGGCGGCGGG
CCGGTATTGTTCTTGCATAGTCATTCTGTTTTCGCTTAAAAACGTTGGAAAAATAAAGTC

GGCATCAATTATAACAGGTTGCCGGAAGCGGCGAATCGGCAGATTGCCGGCAGGATGCGT
AAATTCGCACGCGCATTATTCCGTATGCCGTACAAATACACCGCGTTTATTGATACGCAC
GTTTTTTATGCTAATATTACAAACCAAAATCAAATGTTTAAAACTCTCCTGATGCGGCTC
TTCCGAACAAAAGGCAGACGGGCATCGGGTAAAAGAGGATTCTGCATATGAAAATCAAAC
AAATCGTCAAACCGGGCTTGGCAGTATTGGCGGCGGGCGTTCTGTCTGCCTGCGCAACCA
AAAGCAACGTCAAAGCCGACGGCACGACCGACAATCCGGTTTTCCCGAAACCCTATTCCG
TAACGCTCGACAACAAGCGCGGCACATTCCCGACTTATGACGAACTGGATCAGATGCGCC
CCGGCCTGACCAAAGACGACATCTACAAAATCCTGGGCCGCCCGCATTACGACGAAAGTA
TGTACGGCGTGCGCGAATGGGATTACCTGTTCCACTTCCATACCCCGGGCGTAGGTATCG
ACCCTGAAAACACTTCCGGCGTAGAAGATGTTACTACCTGCCAATACAAAGTGATTTTCG
ATAAAGACAAATTTGCCCGCAGCTTCTACTGGAACCCCGTCTTCCCGAAAGATGCCGCCT
GTCCGCCGCCCGCACCCAAAGCCGAGCCGCAAGTCATCATCCGCGAAATCGTGCCGGCAA
AACCGAAACGTATCCGCCAATAATCCGACATGCCGTTCCGCCTGTTTTTAGGGATATTAT
GCGGCCTGTCAATGGTTGCCCCCGTATATGCACAGGGGCAGCCGGATACGGTCGGCGACT
TTATCCAAAAGAAAAAAGTCATCGTCGATACATCCAAAGCGGAACTCTGTTTCGCTGACG
ACCGTCAGTGCCACCCCGTCCTCATCGGTGTTGCCACGCCCAAGGGGACGTTCGGGCTGA
CGCTGAACAGTACCGACAAGCCCGGATACGGCGGCGAAGTCATCGGTTTCAAGCAGGAGG
GTGATTTTCTTTTCGCCCTGCACCGCGTTTGGAATCAGATACCGTCGGAAAGGCGGAACG
AACGCATCGCCTCCCCGTCCGTGTCCGACAGGATTATGACCAACGGCTGCATCAACGTCA
GCGATGCGGTGTACGAAAAACTGCGTCATTATTTTGTGTTGGAAGTGATTTGAAACAGAC
GGATACCGCACGCGCCGGTATCTGTTTTCACATTGCCCCGATGCCTGAAACAGACTGTCC
GCCACGTCATGCCGTCTGAAACCGGCGCAGATGCCGCCAAGCCTTCAGACGGCATTGCCT
GCCCGCTCCGACCGAACAACAACCATCTTTGGGAGAACCTTATGCCCGAACAAAACCGCA
TCCTCTGCCGCGAACTGAGCTTGCTGGCATTCAACCGCCGCGTGTTGGCGCAGGCGGAAG
ACCAAAACGTCCCCCTTTTGGAACGCCTGCGCTTCCTGTGCATCGTTTCATCCAACCTCG
ACGAGTTTTTCGAAGTCCGTATGGCGTGGCTGAAGCGCGAACACAAACGCTGCCCGCAGC
GCAGGCTGGACAACGGCAAAATGCCGTCTGAAACCATCGCCGACGTTACCGAAGCGGCGC
GCTCCCTGATACGGCACCAGTACGACCTGTTCAACAACGTCCTTCAGCCCGAGCTGGCAC
AAGAAGGCATCCATTTTTACCGCCGCCGAAATTGGACAGACACACAGAAAAAATGGATTG
AAGACTATTTCGACCGCGAATTGCTGCCGATCCTGACCCCCATCGGACTCGACCCTTCCC
ACCCCTTCCCGCGCCCGCTGAACAAATCGCTCAACTTCGCCGTCGAACTCGACGGCACAG
ACGCGTTCGGCAGGCCTTCGGGGATGGCGATTGTGCAGGCACCACGCATCCTGCCGCGCG
TTGTTCCCCTGCCGTCCGAACTGTGTGGCGGCGGACACGGCTTCGTCTTCCTCTCCTCCA
TCCTGCACGCCCACGTCGGAAAACTCTTCCCGGGCATGAACGTCAAAGGCTGCCACCAGT
TCCGCCTGACGCGCGACAGCGACTTGACCGTTGACGAAGAAGACCTGCAAAACCTCCGCG
CCGCCATTCAAAACGAGTTGCACGACCGCGAATACGGCGACGGCGTGCGGCTCGAAGTCG
CCGACACCTGTCCCGCCTACATCCGCGACTTTCTGCTCGCGCAATTCAAACTGACCGCCG
CCGAACTCTATCAGGTCAAAGGCCCGGTCAACCTCGTGCGCCTCAACGCCGTCCCCGACC
TAGTCAACCGCCCCGATTTGAAATTTCCCACACACACGCCGGGCAGACTGAAAGCCTTGG
GCAAAACCGCGTCCATATTCGATTTGGTGCGCCAATCGCCCATCCTGCTGCACCACCCCT
ACCAATCGTTCGACCCCGTTGTCGAAATGATGCGCGAAGCCGCCGCCGACCCCGCCGTGC
TTGCCGTCAAAATGACGATTTACCGCACCGGCACGCGTTCCGAACTCGTCCGCGCCCTGA
TGAAGGCGGCACTCGCCGGCAAACAAGTAACCGTCGTCGTCGAACTGATGGCGCGTTTTG
ACGAAGCCAACAACGTCAACTGGGCGAAGCAGCTCGAAGAGGCGGGCGCGCACGTCGTGT
ACGGCGTGTTCGGCTACAAAGTCCACGCCAAAATGGCACTGGTCATCCGCCGCGAAGACG
GCGTGCTCAAACGTTACGCCCATCTCGGCACGGGCAACTACCACCAAGGCACATCGCGCA
TCTACACCGACTTCGGCCTCATTACCGCCGACGAACAAATCACCGCCGATGTGAACATAT
TGTTTATGGAAATCACAGGTTTGGGCAAACCCGGGCGGCTGAACAAACTCTACCAAAGTC
CGTTTACCCTGCACAAAATGGTTATCGACCGCATCGCACGCGAAACCGAACACGCAAAAG
CCGGCAAACCGGCGCGGATTACCGCCAAGATGAATTCGCTCATCGAACCGACCGTCATCG
AAGCCCTGTATCGGGCAAGCGCGGCAGGCGTACAAATCGATTTGATTGTGCGCGGTATGT
GCACCTTGCGCCCGGGTGTAAAAGGCTTGTCCGAAAACATCCGCGTCCGCTCCATCGTCG
GCAGGCAGCTCGAACACGCGCGCGTGTATTACTTCCATAACAACGGCACGGACGATACCT
TTATCTCCAGCGCGGATTGGATGGGGCGCAACTTCTTCCGCCGCATCGAAACCGCCACGC
CGATTACCGCGCCCGAACTCAAAAAGCGCGTTATACATGAAGGACTGACCATGGCACTGG
ACGACAACACCCACGCGTGGCTGATGCAGCCCGACGGCGGCTATATCCGCGCCGCACCTG
CCGAGGGCGAATCCGAAGCCGACCTGCAAAACGATTTGTGGACACTGCTCGGAGGCTGAC

CCGCACCGCCCCAATCAAAAACCATGCCGTCTGAAACCTTTCCGTTTCAGACGGCATGGT
TTTACAGCAATCTAAACAGGGCGGACCGGAGTCAAAAACACACCTTCGCCATTCCTGCAC
AAGCACTTCCCCTATACGCTCCCAACCCCAAGCCGCCGCATTCCAGACGGCATTATAGTG
GATTAAATTTTAGGGGCTGTACTAGATTAGCAGATATGTTACCCTCGAAATATGAAGATA
ACGCACTGCAAATTAAAGAAAAAAGTACAGAAAGAACTGCTCCGTTTTTTGTGCTGGAAG
TTACCGCCCGTTCTGCCGCCGATATTTTGGGTATCCATCCCAATTCGGCAGCACTGTTCT
ACCGTAAAATCCGCACGGTTATCAACCATCATTTAGCCTTGGCTGCCGATGAGGTTTTTG
AGGGCCCTGTCGAGCCGGACGAAAGCGATTTCGGCGGACGGCGTAAAGGCAGACGTGGTC
GCGGTGCGGCAGGAAAAGTGGTTGTCTTCGGCATTCTGAAACGCAACGGACGGGGCTATA
CCGTTGTCGTAGATAATGCCAAGTCTGAAACGTTACTCCCTGTCATCAAGAAGAAAATCA
TGCCGGACAGCATTGTTTATACCGATAGTCTGAGCAGCTGCGACAAGTTGGACGTGAGCG
GTTTTATCCATTACCGCATCAACCATTCCAAGGAGTTTGCAGACCGTCAGAACCACATTA
ACGGCATTGAGAATTTTTGGAATCAGGCAAAACGCGTCTTGCGAAAATTATAGTGGATTA
ACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAAATAGTACGAAACCGATTCA
CTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCCGTA
CTGGTTTTTGTTCATCCACTATACCTTTCCGACAGCCGAACAAAACCCCGAATCCGTCTG
CACGGTTCGGGGTATATCTCCAATACGGGCATCGTGTTCCGGAAAACCGTCAAATCCGCA
TCGGCATCACAATATATTTGAAATTCGGATTGTTCGGCACGGTAAACAGCGTCGAGCGGT
TGGCATCGCCGAAGGCAAGCTGCATATCGTCGGAATGGATGTTGCGCAACACGTCCATCA
GATAGCCGATATTGAAACCGACTTCGAGTTCGCCGCCCTGATAGGCGATTTCGATTTCTT
CGCGCGCTTCTTCCTGCTCGTTGTTGCTGCACACAACGCTCAACAGGCCGGGTTGCAAAA
ACAATCGCGCACCGCGGAATTTTTCATTGGCAAGAATCGATGCACGTTCCAACGCGCCCA
ACAATTCTGCCCTCGACAACACGAAAATCTTGTCGTTGTCCAAAGGAATCACGCGGTTGA
AATCGGGGAATTTGCCGTCGATGACCTTGCTGACGATGGTCGTGCCGTTGCATTGGAAAC
GCACCTGTTTGTCCAGCAGCTCGATTTGAATCGGATCGTCGGGGTTGTTCAACAGTTTGA
ACAGTTCCAGCACCGTTTTGCGCGGCAAAATCACTTCGGCGCGCGGCAAATCCGCATCAA
TCGCGCAGGCTGCATAGGCAAGGCGGTGTCCGTCGGTCGCCACAAGGCGCAACTGGCTGC
CCTCAACCTGCATCAGCAGACCGTTGAGATAATAGCGGATGTCCTGCACCGCCATGCTGT
ACTGCACTTGCGACAGCATGGTTTTGAAACGCTCCTGCTCCAGCGAGAAAGTCGCGCTGA
TGTCCTCGCCGACATTCATCATCGGAAAATCGGCGGCAGGCAGGGTCTGCAGGGCAAAAC
GCGATTTGCCCGCCTTCAGCGTCAGACGGCTGTCGTCCCAATCCAGCGACACCAGCGCAC
CGGCAGGCAGCGCGCGCAAAATATCCTGAAATTTCTTGGCATTGGTGGTGATGCGGAAGT
CGCCCGCGCCGCCCTCGGGACCCGCAGTGTCGATTTGGATTTCCAAATCGGTTGCCAAGA
GTTTGGTCTGACCGCCTTTTCCCTCAATCAGGACGTTGGACAGGATGGGCAGGGTGTGGC
GGCGTTCGACGATGCCGGTAACGGCTTGCAACGGCTTGAGCAGGCTGTCGCGCTCGGCTT
GTAAAATCAACATGTTCGCTCCTTTAAATCGGTTTGTATAGTGGATTAAATTTAAATCAG
GACAAGGCGACGAAGCCGCAGACGGTACAAATAGTACGGAACCGATTCACTTGGTGCTTC
AGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCCGTACTGGTTTAAAG
TTAATCCGCTATATCTTTACCCTTCGGACGGCATGGGCAATATCATGTCGTCTGAAAACG
TTTTCCATCAGTTTTGAATCAGAATCAGCAGCTTTTCATAATCCTGAGCCAATTCCGGAT
CTTCTTCGCGCAGTTTCGCCACTGCCCTGATGCCGTGCATAACGGTCGTATGGTCGCGCC
CACCAAACGAATCGCCGATAGACGGCAGGCTCAAAGTAGTCAGTTCTTTGGTCAGGCTCA
TCGCCACCTGGCGCGGACGGGCAATGTTTCGTGTCCGTTTCTTACCGAGCACATCGCTGA
TTTTGATGCGGTAATATTTCGCCACCGCATCGATGATGATGTCGGCGGTGATGACTTTGT
GCTTCTCGGCAATAATGTCCTGCAAAGCGGTACGCGCCAAATCGATGTCGATGACGGGAC
GGTTCATAAAGCGGCTGCTCGCTCCGACACGATTAAACGCGCCTTCAAGCTCGCGCACGT
TGGAACGGATCAGATTGGCAATGAACAGCGCGGCTTCGtCTTCGATACTGATGCCCGCCG
CTTCCGCCTTTTTCTGCAAAATGGCGATGCGCATTTCCAATTCGGGCGGCTCGAGTTCCA
AAGTCAGTCCCCATGAAAAACGGGATTTGAGGCGGTCGTCCATGCCTTCGATTTTCGCAG
GCAACACATCGCAAGTGAGGATGAGCTGTTTTTCTCGTTGTGGAAATGGTTGTACAGAT
AGAAAAACTCTTCCATCGTACGGTCTTTGCCTTTGATGAACTGGATGTCGTCGATAATCA
GCAGGTCGTATTGCTTGTATTGCTGCTTGAACACGTCGTAAGTGTTGTTGCGAACCGCCT
TCATAAAGCTGCGGATATAGTCATCCGAATGCATATAGCGCACTTTGGCATCGGGACGGT
TTTTCAGCAGCTCGTTGCCGACCGCCTGCACAAGGTGGGTTTTGCCCAAACCCGTGCTGC
CATAGAGGAAGAACGGGTTGTAACTCTGCCCCGGGCTTTCCGCAATCGCCTGCGCCGCAG
CCGCCGCAAGGCGGTTGCCCTTACCTTCTACCAACGTATCAAACGTGTAATCCGGAGACA
GGTTGGTCTGCTCGTAACGCGCCTCTTCCGCATCGCGCTGCACGTCCGTCCGTGCTTTGG

CAACTGCCACCGATTCCGGCCGGGAAGCAGACCCGGCAGCCTGACGCGGCTCGTGCGGCA
GGTTTTTCATACGTTCCGCCAAAATATCCGCCGCCGTTTTCGACGCAGCGGGTTTGACAG
GCTCTTCAGACGGCAGCTCGTCCAACAGAACCTCCTGCACGGGCATTCCCTCCGACACCG
CATGCAAGGACGGCTCGGCAGGTTCGACAGCACCTTCAACCGCCGCCATCTCATAACGCA
CGCCTTCTCCCGGTTTGAATACGAAGGCGGAACGGCCGGCAGCCAACTCTTCCCTCACCG
CTTCTATTTTTCCGGCAAACTGGCTCTTGAGCATATTGCAGGCAAACTGGTTCTTGCCGT
ACACCACCCATACGCCACCCTCCTCACCAACGGTAAGGGGCGCAATCCATTGCGCAAACT
GCCCTTGAGGCAACATATCGTGAAGACGGCGGAGGCACAGCGGCCAAAACTCTGCTAATG
TCATGGATAGGCTCGAATCGGTAAAAATGAAATCGAAAACAAAGAAAATATAATATTTTC
AAAAGGAAAACAAATCTGTTGAACGCACATCGGTTCAAAACGCGACTGCCCGATTATACC
GACTCACGAATATTTTATCCACAACCCGTGCAAAAATTTATCCACAGAAAGGCGGCGGAA
ATCCGCAGGCAATCGGGCAATCTTCCTGCAAAGTTTCTATATTGATTGACAAAAGCGGCA
AATTGGAGTGTAATTCACGGTTTAATTATCTACCCATTCTATTTTAGGAAACATCATGAA
ACGCACTTATCAACCTTCCGTTACCAAACGCAAACGCACCCACGGCTTCCCTGGTGCGCT
CCAAAACGCGCGGCGGCCGCGCAGTATTGGCCGCACGCCGTGCCAAAGGCCGCAAACGCC
TGGCGGTATAATTTTGGACTACCGCTTCGGAAGGCAGTACCGCTTGTTGAAAACGGATGA
TTTTTCATCCGTTTTTGCATTCAGAAACCGCCGCAGCCGCGACCTGCTGCAAGTTTCGCG
CTCAAACGGCAACGGGCTGGGCCATCCCCGCATCGGTCTGGTGGTCGGCAAAAAAACCGC
CAAACGCGCCAACGAACGAAATTATATGAAGCGCGTTATCCGCGACTGGTTTAGATTGAA
CAAAAACCGGCTGCCGCCGCAGGATTTCGTCGTGCGCGTCCACCGTAAATTCGACAGGGC
TACCGCAAAACAGGCAAGGGCGGAACTGGCACAACTCATGTTCGGCAACCCCGCAACCGG
ATGCAGGAAACAGGCATGATCAGAACGGTACTCTGCAGGCAAGGTTCAGACGGCAACGGG
TTTCCCATACAAGGAACATCCCGATGAACTTCCTATTGTCCAAACTCCTGCTGGGACTGA
TACGGTTCTACCAATATTGCATCAGCCCGCTGATTCCGCCGCGCTGCCGTTATACGCCGA
CCTGTTCGCAATACGCGGTCGAAGCGGTCAAAAAATACGGCGCATTCAAAGGCGGCCGGC
TCGCCATCAAGCGCATTGCACGCTGCCACCCTTTCGGCGGACACGGACACGACCCCGTTC
CCTGACCCGACGCAATATTCAAATTGCACGCTTTCCTTTTATTTCCCATCGGTTTCTATA
TAATGCCGTCTGAAGCTTCGGGCAGGCGGCACGACCGCCGGGTATGAAGCCCGCCCTTAT
TCCCCGTCTATCGGAACACGCAACCTGCGGCATTTCCGACCATTCAkGAAACTCTTATGG
ATTTTAAAAGACTCACGGCGTTTTTCGCCATCGCGCTGGTGATTATGATCGGCTGGGAAA
AGATGTTCCCCACTCCGAAGCCCGTCCCCGCGCCCCAACAGGCAGCACAACAACAGGCCG
TAACCGCTTCCGCCGAAGCCGCGCTCGCGCCCGCAACGCCGATTACCGTAACGACCGACA
CGGTTCAAGCCGTCATTGATGAAAAAAGCGGCGACCTGCGCCGGCTGACCCTGCTCAAAT
ACAAAGCAACCGGCGACGAAAATAAACCGTTCATCCTGTTTGGCGACGGCAAAGAATACA
CCTACGTCGCCCAATCCGAACTTTTGGACGCGCAGGGCAACAACATTCTAAAAGGCATCG
GCTTTAGCGCACCGAAAAAACAGTACAGCTTGGAAGGCGACAAAGTTGAAGTCCGCCTGA
GCGCGCCTGAAACACGCGGTCTGAAAATCGACAAAGTTTATACTTTCACCAAAGGCAGCT
ATCTGGTCAACGTCCGCTTCGACATCGCCAACGGCAGCGGTCAAACCGCCAACCTGAGCG
CGGACTACCGCATCGTCCGCGACCACAGCGAACCCGAGGGTCAAGGTTACTTTACCCACT
CTTACGTCGGCCCTGTTGTTTATACCCCTGAAGGCAACTTCCAAAAAGTCAGCTTTTCCG
ACTTGGACGACGATGCCAAATCCGGCAAATCCGAGGCCGAATACATCCGCAAAACCCCGA
CCGGCTGGCTCGGCATGATTGAACACCACTTCATGTCCACCTGGATTCTCCAACCTAAAG
GCAGACAAAGCGTTTGCGCCGCAGGCGAGTGCAACATCGACATCAAACGCCGCAACGACA
AGCTGTACAGCACCAGCGTCAGCGTGCCTTTAGCCGCCATCCAAAACGGCGCGAAAGCCG
AAGCCTCCATCAACCTCTACGCCGGCCCGCAGACCACATCCGTCATCGCAAACATCGCCG
ACAACCTGCAACTGGCCAAAGACTACGGCAAAGTACACTGGTTCGCCTCCCCGCTCTTCT
GGCTCCTGAACCAACTGCACAACATCATCGGCAACTGGGGCTGGGCGATTATCGTTTTAA
CCATCATCGTCAAAGCCGTACTGTATCCATTGACCAACGCCTCTTACCGCTCTATGGCGA
AAATGCGTGCCGCCGCACCCAAACTGCAAGCCATCAAAGAGAAATACGGCGACGACCGTA
TGGCGCAACAACAGGCGATGATGCAGCTTTACACAGACGAGAAAATCAACCCGCTGGGCG
GCTGCCTGCCTATGCTGTTGCAAATCCCCGTCTTCATCGGATTGTATTGGGCATTGTTCG
CCTCCGTAGAATTGCGCCAGGCACCTTGGCTGGGTTGGATTACCGACCTCAGCCGCGCCG
ACCCCTACTACATCCTGCCCATCATTATGGCGGCAACGATGTTCGCCCAAACTTATCTGA
ACCCGCCGCCGACCGACCCGATGCAGGCGAAAATGATGAAAATCATGCCGTTGGTTTTCT
CCGTCATGTTCTTCTTCTTCCCTGCCGGTCTGGTATTGTACTGGGTAGTCAACAACCTCC
TGACCATCGCCCAGCAATGGCACATCAACCGCAGCATCGAAAAACAACGCGCCCAAGGCG
AAGTCGTTTCCTAAATGCCGCAGCATGAAAAATGCCGTCTGAAACCTGTTCAGACGGCAT

TTTTATTGCCCACCCCCTATCGGGGCGGAAATCTTCAACCCGCATACATCACAAAAATCG
TCGGGCGTTTTTTCAGATTGGGCATTTCTTTTCTTTTTCGCCACTGCACGATTGTTTGAC
TGATGATTTCCTGTGTCGGCAAGGTCAAATCCGTAGCCGTGCATAAACGCGTTTCAGGAT
GCAGGTTTTCCACCGCATCGGCAAGCAGCGCATCATTGCGGTAAGGCGTTTCAATAAAAA
TCTGCGTCTCGCCGCACTGGCGCGAACGCTGTTCCAAAGCCCGAAAAGCCTGAATCCGCT
CGTTTTTTTCAGACGGCAGATAGCCTTTAAACGCAAAACTCTGCCCGTTCGCACCCGAAG
CCATCAAAGCCAGCAGCAGGCTGGAAGGCCCGACCAGCGGACGCACTTCAAAACCGTGTT
TATGCGCCAATGCCACCAAATTCGCACCCGGATCGGCCACAGCCGGGCAACCCGCCTCAC
TGACAATGCCCATACTGCGCCCTTCTTGCAAAGGTTTCAGCAATTCCGGCAAAGTCTTCA
AATCCGTATGTTCATTCAACGTTTGCAGATTCAGCTCGCGGATAGGCGTAGTCACGCCCA
AATGTTTCAAATGCGCACGCGCCGTTTTTTCCGCCTCCACGACAAAATCCGTCAGCCCGA
CAATCGCCTGTTGTTCATGCGGCAACAGGCACGGCGTGTCAGGCGTACCCAAAGGCGTAG
GAATCAAATACAAAACAGGAGACATCATTCCCTCACTCATCGGTTAAAAATGCCGTCTGA
GCCTTTCAGACGGCATAAACGGGCAGTTACAGAACCTCCACGCCCTCATTTTTCAAGAAA
TCGACCAGACGGAAAACCGGCAAACCGATTAAAGCATTCGGATCGGTACTCTCAATCCTT
TCAATCAGCAATGCACCCAAATCCTCACTCTTCAGCGCACACGAACAATAAACCGCATCA
GGCTCGCGCTCCAAATAGCGGAGGATATGCAACTCGTCCAACTGCCTCATCACGACCACC
GTCTTATCGATATGCCGCCGCATCCTGCCCGTAACCGTATTCAACAGCACGATCGCGCTG
TAAAACTCAATCTCCCTGCCGCTCAAGTGCATCAGCATCTTTTGCGCATTGGCAAGGTTC
ATCGGCTTGCCCCACTGCCTGCCGTCGCACCACGCCACCTGGTCCGCACCGACAATCAAC
GCCTCTGGGAAACGCCCGGTCAACGACCGCGCCTTACCCTCGGCAAGGCGCAATGCCGTC
TGAGGGGCGGATTCCCCCAACATCGGCGTTTCGTCAAAATCGGGGGACGCCGCCTGAAAG
GCAATGCCGAGCCTTTCCATCTGTTCGCGGCGGAAAACCGAACTCGTACCCAAAATCAAA
GGCAGTTCCAAACCCATCCCATCCTCCTTACCGTTGAAAACACGCCCGAAGGGGCAGTAA
AATCCAGCCATGCGCCGAAACACGGATACCCGCCTTCGGCGTACCGCAACATTTTTCTTA
AAAATATTGACGTTAGAACATCTAAATTATATCATATCCCGTTTATGTCAGACCCTAATT
TGATTGACTTGGAAATTTTTGCCGCCGAAGGGCAGAACCTGCAAGGCAGTTTCTGCTGG
AAGAATTGGATGAACGCGTCAGTTCGCACGATTATCCCGCCGACAGGCAGACCAAAATAT
CGTTTACACTGACCGGCGGTCGCGACCGGCTGCAACGCCTGTTCCTCGACCTGAACGTCA
AAGCCGATATGCCCCTGATTTGCCAGAGATGTATCAAACCCATGCCGTTCATGCTTGATG
AAAGCAGCCGTATCGTCCTGTTTTCCAACGAAGAGTCCTTGGACGAATCCATGCTTGCCG
ACGAAGAACTCGAAGGCATACTGATTGAAAAAGAACTCGACGTGCGCACATTGGTAGAAG
ACCAAATCCTGATGTCCCTGCCCTTTTCGCCGCGACACGAAGACTGCGGCGACAATGGGA
CACTGGAAGAAGTCAATCGGGACAAACCCAACCCCTTTGCTGTTTTGGCAGGTTTGAAAA
GCAATTGATTAGGACACAGTTTATTTATCTAGGAGCTTGAAATGGCCGTTCAACAAAACA
AAAAATCCCCTTCCAAACGCGGTATGCACCGTTCGCACGACGCGCTGACCGCGCCTGCAC
TGTCTGTCGACAGCACAACCGGCGAAGTACACCGCCCGCACCACATCTCCCCCAACGGTA
TGTACCGCGGCCGCAAAGTGGTCAAAGCCAAAGGCGAATAATCCCTATTCGACTGACTGA
AAAAGCCAGAACATTGCCATGCAATTACTGGCTTTTTTTGCATTGGACGCACCATCCGTC
CAAACTTTCGCCATACGTCAACACACAGGGGCAAAGCGTTCCGTATAATACCCCGTGAAA
ATATTCCAAAAGCCCCAACCACCAAGGAAATTCCGATGAAACAGAAAATCTGGTACACCT
ACGATGACATCCACCGCGTCATCAAAGCATTGGCAGAAAAAATCCGGAACGCCGACATCA
AATACGATGCCATGATTGCCATCGGCGGCGGCGGCTTTATTCCGGCACGTATGCTGCGCT
GTTTTCTGGAAATTCCGATTTATGCCGTAACCACCGCCTATTACGACAGCGACAACGAAG
GACAGGTTACCGAAGAAGTCAAAAAAGTCCAATGGCTCGACCCCGTTCCCGAAGCCCTGC
GGGGCAAAAACGTACTCGTCGTCGATGAAGTGGACGACAGCCGCGTAACCATGGAGTTCT
GCCTGAAAGAACTGCTCAAGGAAGACTTCGGTACGATCGGAGTCGCCGTACTGCACGAAA
AAATCAAAGCCAAAGCAGGCAAAATCCCCGAAGGCATTCCCTATTTCAGCGGCATCACCG
TAGAAGACTGGTGGATCAACTATCCGTGGGACGCACTCGACATCGACGAACACAACCGCC
TTGCCCAGGCCGGCCGAGGCTGACCCTTTCAGACGGCATATTTCCGAACCGATGCCGTC
TGAAGCCCGCACGACCCCTGCCGCAGACCGAAAACCTACCGGAGAAACCCTATGATTACA
TTGGCCGTAGATGCCATGGGCGGCGACCAAGGACTTGCCGTTACCGTACCCGGCGCAACC
GCATTCCTCCAAGCACACCCCGATGTCCGCCTGATTATGACCGGCGACGAAACGCAACTG
CGCCAAGCCCTGACCGCGGCAGGCGCACCGATGGAACGCATCGACATCTGCCATACCACC
CAAGTCGTCGGCATGGACGAAGCCCCGCAATCCGCCCTGAAAAACAAAAAAGACTCCTCC
ATGCGCGTCGCCATCAACCAGGTTAAAGAAGGCAAAGCCCAAGCCGCCGTATCCGCAGGC
AACACGGGTGCGCTCATGGCAACCGCACGTTTCGTCCTCAAAACCATTCCCGGCATCGAA

CGCCCCGCCATCGCCAAATTCCTTCCTTCCGACACCGACCACGTTACCCTTGCACTCGAC
CTTGGCGCGAACGTCGACTGCACGTCCGAACAGCTCGCCCAATTTGCCGTTATCGGCAGC
GAACTCGTCCACGCACTCCATCCTCAAAAAGGACAGCCGCGCGTCGGGCTGGTCAACGTC
GGCACGGAAGACATCAAAGGTACGGACACCGTCAAACAAACCTACAAACTGCTGCAAAAC
AGCAAACTCAACTTTATCGGCAACATCGAAAGCAACGGCATCCTCTACGGCGAAGCAGAT
GTCGTCGTCGCCGACGGCTTTGTCGGCAACGTCATGCTCAAAACCATCGAAGGCGCGGTC
AAATTCATGAGCGGAGCCATCCGCCGCGAATTCCAAAGCAACCTGTTCAACAAACTTGCC
GCCGTTGCCGCCCTACCCGCCCTCAAAGGGCTGAAAAACAAACTCGACCCGCGCAAATTC
AACGGGGCCATCCTGCTCGGGCTGCGCGGCATCGTGATTAAAAGCCACGGCGGCACAGAC
GAAACCGGTTTCCGCTATGCCCTCGAAGAAGCCTACCACGAAGCCAAGTCCGCCGGCCTT
TCCAAAATCGAACAGGGCGTAGCCGAACAACTCGCCGCACTCGAAACTGCCAAAGCCGTC
CAAAACGAAAATGTCGGCGGTCTGTAACACACACGATGCCGTCTGAACGCCCCCGCCCCT
TTCAGACGGCATCCGCCCGCACCAAACCTGCGGGCGCGGACGGCGATGCGCCTGTCCGGC
ACTTCCCAAATATCGCCTTGTAAAATAAGGAGTATTTGAAAAATGAAGACATTAGAAAAA
CGGATGAAAGCTCTAGACAAACGGATTATGAAGTTCGGAAAATCCCTTGAAGGCAGGCTT
GATGCCCGTCTGATTGAATCCGCATTGGATTATATTCATTATTCGGAACGTTTTTTGGCT
TTTGAAATCCTGTGTACTTATATCGAAGATTTCGATGTCCGGCTGACGGAACAAGAATCC
CGGGAAATTTCTTTTATCAACAAGGAATTTGAGATAGAAAGCACGTCCGATTAACCAATA
AAGCCAATGGGTTGATAAACATGAAAACATCGACGGTCGTTTTTGGCGGATTTTTTATGG
CAGACAACGGAGAGCGAATCCAAATCCCCGTTTTGGAAAATCCTGACATTAGGGAAATCA
ATCACTTTTTTTCCGTATCAAATTTTGAGAAAAAAACCGGCGTCCTTGTTTTCAGAATCA
TCCCCGAGCCGGAATTTGGCAATACCGAATTAACTGTCTATTTTAAAAAAGGATATTATA
GTGGATTAACAAAAACCAGTACGGCGTTGCCTCGCCTTGCCGTACTGGTTTTTGTTAATC
CACTATATCAGACGAAAACAAACACCCGCGCCAATAGCCTGACGGCAACCCGGCAATCAA
AATGCCGTCTGAAGCAGCTTGGGCTTTCAGACGGCATTTCCTTCGCTTAAAACAGCGTAT
CGGCAACCCCGCCCTGCCTGTCCACGGCAATCTGCATCTGAAAACCATCTGTATCCCAAA
CCACACCCCCATCCCTGTTTCCATCATGTGCACCCTGTCCGTATTGGGCAATCATCTGTT
TTTCGCTTACAATAGCCGAATCTGAACCAACTCTCTAAAAAGGCCGTTCCCATGCAGTAT
GCAAAAATTTCCGGCACAGGCAGCTATCTTCCCGCCAACCGCGTCAGCAATGACGACCTT
GCCCAAAAGGTAGATACCTCTGACGAGTGGATTACCGCGCGCACGGGCATCAAATTCCGC
CATATTGCAGCCGAAAACGAAAAAACCAGCGATCTTGCCGCCGAAGCGGCGCACCGCGCG
CTGGATGCAGCCGGATTAGACAGCGGCGAAATCGATTTGATTATCGTGGCAACGGCAACG
CCGGATATGCAGTTTCCGTCTACTGCGACCATCGTGCAACAAAAATTGGGCATCACCAAC
GGCTGCCCCGCGTTTGACGTACAGGCGGTGTGCGCCGGCTTTATGTACGCGCTGACCACG
GCAAACGCCTACATTAAAAGCGGTATGGCGAAAAACGCGCTGGTCATCGGCGCGGAAACC
TTCAGCCGCATTGTAGACTGGAACGACCGCACAACCTGCGTATTGTTCGGCGACGGCGCG
GGCGCGGTGGTTTTAAGCGCGTCGGACACGCCGGGCATCATCCACAGCAAACTCAAGGCC
GACGGCAATTATCTGAAACTCTTAAACGTCCCCGGGCAAATCGCCTGCGGCAAAGTTTCC
GGTTCGCCGTACATTTCGATGGACGGTCCCGGCGTGTTCAAGTTTGCCGTCAAAATGCTG
TCCAAAATCGCCGATGACGTTATCGAAGAAGCAGGTTACACCGCCGCTCAAATCGACTGG
ATTGTTCCCCATCAGGCAAACCGCCGCATTATCGAATCGACCGCGAAACATTTAGGTTTG
AGTATGGACAAAGTCGTCCTGACCGTCCAAGACCACGGCAACACATCCGCCGCATCGATT
CCGCTGGCTTTGGATACGGGCATCCGCAGCGGACAAATCAAACGCGGTCAAAACCTGCTG
CTCGAAGGCATCGGCGGCGGTTTCGCGTGGGGCGCGGTGCTGTTGCAATATTGAACCCGA
TGCCGTCTGAAACAGGCTTTCAGACGGCATTTCCCATATCATGAAGCGGCAGGCTTTCTT
CAAACTGATGGCGTGTGCGGCATTTCTGTCTGCCGTTTCGCTGCGCCTCCCCGTATTGGG
CGCGTGTTACGCAATATTGTCCCTCTATGCGTTTGCACTTTACGGCATCGACAAACGGTG
CGCCATACGGGGGCAACGCCGCATTCCCGAACACCGCCTGCTGCTGCCTGCATTGCTCGG
CGGCTGGGTGGGCGCGTATTTCGGCAGCATGACATTCAAACATAAGACAGCGAAAAAGCG
TTTTGTTCTGCTGTTCCGTCTGACTGTTTCAGGTAATGTCTTGGCGACCCTCATCCTGAT
TTATAGTGGATTAAATTTAAACCAGTACGGCGTTGCCTCGCCTTGCCGTACTATTTGTAC
TGTCTGCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATATTATTTTGTCCCG
CCTGAATTTTTCGTAAAACTCGGGCAGAATACCTGATTATCCAACCAAACAAAGGAATAC
TATGTCTTTTGCCTTCTTTTTTCCCGGACAAGGTTCCCAAAGCCTCGGTATGATGAACGG
CTTTGCCGAACACGCCATCGTCAAAAACACCTTTGCCGAAGCCTCCGCCATATTGGGGCA
GGACTTGTGGGCGATGATAAACGGCAGCGATGCCGAAATCATCGGTCAAACCGTCAACAC
CCAGCCCATTATGCTCGCCGCCGGCGTTGCCGTTTACCGCGCCTATTTAGAAGCGGGCGG

CAAAACGCCTGCCGCCGTTGCCGGACACAGCCTCGGCGAATACACCGCACTCGTTGCCGC
CGGCGCATTGAATTTTGCCGACGCGGTCAAACTCGTGCGCCTGCGCGCCGAACTGATGCA
GTCCGCCGTACCGCAAGGCGTGGGCGCAATGGCGGCGATTCTCGGCTTGGAAGATGAGCA
GGTTAAAGCCATTTGTGCCGAAGCCGCCCAAAGCGAAGTGGTCGAAGCCGTCAACTTCAA
CTCACCCGGACAAATCGTGATTGCAGGCAACGCCGCCGCCGTCGGACGCGCCATGGCTGC
CGCCAAAGAAGCCGGTGCCAAACGCGCCCTGCCGCTGCCCGTGTCCGTACCTTCCCATTG
CAGCCTGATGAAACCCGCCGCCGACAAACTTGCCGAAGCCCTGAAAACCGTTGAAATCAA
GCAGCCGCAAATCCGCGTTATCCACAACGCCGACGTTGCCGCCTACGATGATGCCGACAA
AATCAAAGACGCGCTCGTCCGCCAGCTTTACAGCCCCGTACGCTGGACGGAAACCGTCAA
CGCCCTCGTTTCAGACGGCATTGCCGAATCCGCCGAATGCGGCCCGGGCAAAGTGTTGGC
GGGCTTGGCAAAACGCATCAACAAAGCCGCCGCGTGCAGCGCACTGACCGATGCCGGACA
GGTTGCCGCCTTTATCGAAGCGCACTGACTTCGTTCTGCAAAAAGCAGCCTGCCCTCTTC
AGGCTGCTTTTCATGTCCGAACGACGGCAGCCCCATATTTACGCTATAATCCATCCCGAC
CAAACCACCGACAGCGGCTGCCGTTGCAGTTCCCGCCCTACCGATATGATAGAAAAACTG
ACTTTCGGACTGTTTAAAAAAGAAGACGCGCGCAGCTTTATGCGCCTGATGGCGTACGTC
CGCCCCTACAAAATCCGCATCGTTGCCGCCCTGATTGCCATTTTCGGCGTTGCCGCCACC
GAAAGCTACCTTGCCGCCTTCATCGCCCCCCTGATTAACCACGGCTTTTCCGCACCTGCC
GCGCCGCCCGAGCTGTCTGCCGCCGCCGGCATCATTTCCACCCTGCAAAACTGGCGCGAA
CAGTTTACCTATATGGTTTGGGGGACGGAAAACAAAATCTGGACCGTCCCGCTCTTCCTC
ATCATCCTCGTCGTCATCCGTGGCATCTGCCGCTTTACCAGCACCTATCTGATGACTTGG
GTCTCCGTGATGACCATCAGCAAAATCCGCAAAGATATGTTTGCCAAAATGCTGACCCTT
TCCTCCCGCTACCATCAGGAAACGCCGTCCGGCACCGTACTGATGAATATGCTCAACCTG
ACCGAACAGTCGGTCAGCAACGCCAGCGACATCTTCACCGTCCTCACGCGCGACACGATG
ATCGTTACCGGCCTGACCATCGTCCTGCTTTACCTCAACTGGCAGCTCAGCCTCATCGTC
GTCCTGATGTTCCCCCTGCTCTCCCTGCTCTCGCGCTACTACCGCGACCGTCTGAAACAC
GTCATTTCCGACTCGCAAAAAAGCATAGGCACGATGAACAACGTGATTGCCGAAACCCAT
CAGGGACACCGCGTCGTCAAGCTGTTCAACGGGCAGGCGCAGGCGGCAAACCGGTTCGAC
GCGGTCAACCGCACCATCGTCCGCCTCAGCAAAAAAATCACGCAGGCAACGGCGGCACAT
TCCCCGTTCAGCGAACTGATCGCCTCGATCGCCCTCGCCGTCGTCATCTTCATCGCCCTG
TGGCAAAGCCAAAACGGCTACACCACCATCGGCGAATTTATGGCATTCATCGTCGCGATG
CTGCAAATGTACGCCCCCATCAAAAGCCTTGCCAACATCAGCATCCCTATGCAGACGATG
TTCCTCGCCGCCGACGGTGTATGTGCATTTCTCGACACCCCGCCCGAACAGGACAAGGGC
ACGCTCGCACCGCAGCGTGTCGAAGGGCGCATCAGCTTCCGCAACGTCGATGTCGAATAC
CGTTCAGACGGCATCAAAGCCCTCGACAACTTCAACCTCGACATCAGACAAGGCGAACGC
GTCGCCCTGGTCGGACGTTCCGGCAGCGGCAAATCCACCGTCGTCAACCTGCTGCCCCGC
TTTGTCGAACCGTCTGCCGGCAACATCTGCATAGACGGTATCGACATCGCCGACATCAAA
CTCGACTGCCTGCGCGCCCAATTCGCCCTCGTCTCCCAAGACGTATTCCTGTTTGACGAC
ACCCTGTTTGAAAACGTCCGATACAGCCGTCCCGACGCGGGCGAAGCCGAAGTCCTGTTC
GCCCTCCAAACCGCCAACCTGCAAAGCCTGATTGACAGCTCCCCGCTCGGACTGCACCAG
CCCATCGGATCGAACGGCAGCAACTTATCCGGCGGACAGCGGCAACGCGTCGCCATTGCC
CGCGCCATTTTGAAAGACGCGCCGATATTATTATTGGACGAAGCCACCAGCGCATTAGAC
AACGAATCCGAACGCCTCGTCCAACAGGCGCTCGAACGCCTGATGGAAAACCGCACCGGC
ATCATCGTCGCCCACCGCCTGACCACCATCGAAGGGGCCGACCGCATCATCGTGATGGAC
GACGGCAAAATCATCGAACAAGGCACACACGAACAACTGATGTCCCAAAACGGTTACTAC
ACGATGTTACGCAATATCTCAAACAAAGATGCCGCCGTCCGGACGGCATAAACAAAATGC
CGTCCGAAATGGTACAATCGCCCCGACCCTTTCAGACGGCATCATATCCGCCGACCCATC
CGATTATCTTCAATCACTGTAAAACCCATTATGACCCAAGACAAAATCCTCATCCTTGAC
TTCGGTTCGCAAGTTACCCAGCTCATCGCCCGCCGCGTGCGCGAAGCCCACGTTTACTGC
GAGCTGCATTCTTTCGATATGCCTTTGGACGAAATCAAAGCCTTCAACCCCAAAGGCATC
ATCCTCTCCGGCGGCCCCAATTCCGTTTACGAATCCGACTATCAAGCCGATACCGGTATT
TTTGATTTGGGCATTCCGGTTTTGGGCATCTGTTACGGCATGCAGTTTATGGCGCACCAC
TTGGGCGGCGAAGTGCAGCCCGGCAACCAGCGCGAATTCGGTTATGCGCAAGTTAAAACC
ATAGACAGCGAGCTGACACGCGGCATTCAAGATGGTGAGCCAAACACACTCGACGTATGG
ATGAGCCACGGCGACAAAGTGTCCAAACTGCCCGACGGTTTCGCCGTCATCGGCAACACC
CCGTCCTGCCCGATTGCCATGATGGAAAACGCCGAAAAACAATTCTACGGCATCCAGTTC
CACCCCGAAGTTACCCACACCAAACAAGGCCGCGCCCTGTTGAACCGCTTTGTCTTGGAT
ATTTGCGGCGCACAACCGGGCTGGACGATGCCGAACTACATCGAAGAAGCCGTTGCCAAA

ATCCGCGAACAGGTCGGCAGCGACGAAGTGATTTTAGGTCTGTCCGGCGGCGTGGACTCT
TCCGTAGCCGCCGCGCTGATTCACCGCGCCATCGGCGACCAACTGACCTGCGTGTTCGTC
GATCACGGTTTGTTGCGCCTGAACGAAAGCAAAATGGTGATGGATATGTTCGCCCGCAAC
TTGGGTGTGAAAGTGATACACGTCGATGCCGAAGGGCAGTTTATGGCGAAACTCGCCGGC
GTAACCGACCCCGAGAAAAAACGCAAAATCATCGGTGCGGAATTTATCGAAGTATTTGAT
GCCGAAGAAAAAAAACTTACCAACGCCAAATGGTTGGCACAAGGCACGATTTACCCTGAC
GTAATCGAATCCGCAGGTGCAAAAACCAAAAAAGCCCACGCCATCAAATCGCACCACAAC
GTCGGCGGCCTGCCTGAAAACATGAAGCTCAAATTGCTTGAGCCTTTGCGCGATTTGTTC
AAAGACGAAGTACGCGAATTGGGTGTGGCTTTGGGCCTGCCGCGCGAAATGGTGTACCGT
CATCCGTTCCCGGGTCCGGGTTTGGGCGTGCGTATTTTGGGCGAAGTGAAAAAAGAATAT
GCCGACCTGCTTCGTCAGGCAGACGATATTTTCATTCAAGAATTGCGCAATACTACCGAT
GAAAACGGTACATCTTGGTACGACCTGACCAGCCAGGCATTCGCCGTGTTCCTGCCCGTC
AAATCTGTCGGCGTAATGGGCGACGGCCGCACATACGATTACGTCATTGCCTTGCGTGCC
GTGATTACCAGCGACTTTATGACCGCGCATTGGGCGGAACTGCCGTATTCCTTGTTGGGC
AAAGTGTCCAACCGCATCATCAACGAAGTCAAAGGCATCAACCGCGTGGTTTATGATGTG
AGCGGCAAACCGCCTGCCACCATCGAGTGGGAATAAACAGCAAACATGGCTGCCCCGTCC
GGCGCAGTCCTTCGATTATCGGAAAAAAGGAAAAAATATGAGCACACAAGATTTAAACGG
CAAAATCGCTTTGGTAACAGGCGCATCGCGCGGTATCGGTGCAGCAATTGCCGACACGCT
GGCGGCAGCCGGTGCCAAAGTCATCGGTACGGCGACCAGTGAGAGCGGTGCGGCGGCGAT
TAGCGAGCGGTTGGCGCAATGGGGCGGCGAAGGCCGCGTATTAAATTCCGCCGAACCTGA
AACCATCGAAAGCCTGATTGCCGACATCGAAAAAGCGTTCGGCAAACTCGATATTCTGGT
CAACAACGCCGGCATCACCCGCGACAACCTCCTGATGCGCATGAAAGAAGAAGAGTGGGA
CGACATCATGCAGGTCAACCTCAAATCCGTGTTCCGCGCTTCTAAAGCCGTTTTGCGCGG
TATGATGAAACAACGTTCCGGCCGCATCATCAACATCACATCCGTCGTCGGCGTGATGGG
CAATGCCGGTCAAACCAACTATGCCGCGGCAAAAGCAGGCTTAATCGGTTTCTCCAAATC
CATGGCGCGCGAAGTCGGCAGCCGGGGCATTACCGTCAACTGCGTCGCCCCTGGCTTTAT
CGATACCGACATGACACGCGCCCTGCCGGAAGAAACCCGCCAAACCTTTACCGCCCAAAC
CGCCTTGGGCAGATTCGGCGACGCGCAAGACATCGCCGATGCGGTTCTGTTCCTCGCTTC
CGACCAAGCAAAATACATCACCGGCCAAACGCTGCACGTCAACGGCGGTATGCTGATGCC
TTAACAGACAACTTTTTCAACCATGCCGTCTGAAGCCCTTTCAGACGGCATTTGCATTCT
CAGGCAAAATGAACACACACCACACCCCGCCCTGCCCATGCGGCTCAGGCACAAGCTGAG
ACCTTTGCAAAATTCCTTTCCCTCCCGACAGCCGAAACCCCAACACAGGTTTTCAGCTGT
TTTCAGCTGTTTTCGCCCCAAATACCGCCTAATTCTACCCAAATACCCCCTTAATCCTCC
CCGGACACCTGATAATCAGGCATCCGGGTCACCTTTTAGGCGGCAGCGGGCGCACTTAGC
CTGTTGGCGGCTTTCAAAAGGTTCAAACACATCGCCTTCAGATGGCTTTGCGCACTCACT
TTAATCAGTCCGAAATAGGCTGCCCGGGCGTAGCGGAATTTATGGTGCAGCGTACCGAAG
CTCTGTTCGACCACATATAGTGGATTAACAAAAACCAGTACGGCGTTGCCTCGCCTTAGC
TCAAAGAGAACGATTCTCTAAGGTGCTGAAGCACCGAGTGAATCGGTTCCGTACTATTTG
TACTGTCTTCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATACATCATCGCT
ACTACCGTTCCGGCGCAACAGGCATTCCTCGATGCCGCCGAACTGATGCAATGGAGTATA
GAAACCGAAGGGCTGGGCTTGAACGTCATCTCGCACAAGATACTCGGCAAAGACCACGCC
CAAGTCGAATTTGAAGCCTACTTCCGAGACGGACAACACCGATCCGCGCATCACGAACTG
TCCGGCTTCGTCAACATCGGCGGACAATGGTATTTTATCGATCCCACCGTTCCGCATCCT
GCGATGAAACAACCCTGCATTTGCGGATCAGGCAAAAAATTCAAAGCCTGCTGCGGCAAA
TATCTGAAACCTGTCGCATAAAAATGCCGTCTGAACGTTCAGACGGCATTTTCAACGTGC
AAAAAAAACCATTCATACCAAGGGTAAGTATGAATGGTCAATACATTGCGGGAAAACGTC
TTACTTGCTGCACTGCCGAAAAGGGAGAAACGGCAGCGGTAATCAGCGGAAAGGATTGTA
CCCGAATTAATATTAAGAAACGTTAATCGCGAAAATATATTAACAAACCTGTTGAAACCT
ATTGGTTTTCCCGTATCCACCCGACCCAGCGTTCAAACAGCTTCGGTTCGAGCGCGGCAA
CGACCGAGCGTTTGAACACGTGTTCACCACTCCAAAACCCGTCGCCTTCCAAAGTCGTCA
GCCTGCCGCCCGCCTCCTCGAAAATCAACGCGCCGGCGGCATAATCCCACAGCTTCTGCC
CGCCGTGAACATAAACATCATAACGCCCGCACGCCAGATAACACCAATCCAACGTACTGC
TGCCCATACTCCGTATCGTTCCAAAAGGCGCGAGCGTACTCATACGGCTGGAAAGTTTGC
CCGAACGCAGATATTTGATTTCCACGCCCGCAATCGCCTCATTGAGTTTTTTATCCACGA
GGCGCAGGGGCAGACGCGTCCCGTTTAAAAACGCCCCCTGCCCGCGTTCGGCATAAAAAC
ATTCGCCGCTGACTGGGTTGTAGATTACGCCCAACTCGGCGCGCCCGTTGCGGACAAACG
CCACCGATACCGCAAAATGCGGCAGCCCGTTGACAAAATTGTTCGTCCCGTCTATCGGAT

CGACAATCCACAGCCCCTTTTCCCCCGAATATTGTTCCCACAAAGCCGACTGTTCCTGCC
GCGACATTTCCTCACCCAACATCGGACTGTCGATTAAAAGCGGCAACGCGGCGGCAAAAG
CCGTCTGCGCGGCAATGTCCGCCTCGCTCAACATCGAACCGTCTTCCTTGCGGTGAGACG
GCGTATTCAAAAAACGCGGCATAATTTCGGTTTGCGCGATATGGCGCACGACTTTCTGCA
AACGGTGTAACACTTCCTACTGTCCTCATATTTTGAACTTGCGGCGCGCGAACGTATAAT
GTCCGCTTCCATCACGCCGCTGCGACGGATTATAACCGTCCGAACCGCCAAAAACTATGC
CCCGATTCCACCTGCCCGAAAACCTTTCCGTCGGACAAACCGTCGCCCTGCCCGACAACA
TCGTCCGCCACCTCAACGTCCTGCGCGTCCGCCCCAACGAAAACATCACCCTCTTCGACG
GCAAAGGCAAGGCACACGCCGCACGGCTGACCGTTTTGGAAAAACGCCGCGCCGAAGCCG
AAATCCTGCACGAAGACACAACCGACAACGAGTCCCCGCTCAACATCACACTGATACAAT
CCATCTCCTCCGGCGATCGCATGGATTTCACCCTGCAAAAAAGCGTCGAACTCGGCGTAA
CCGCCATACAGCCCGTCATCAGCGAACGCTGCATCGTCCGCCTCGATGGGGAACGCGCCG
CCAAACGCCTCGCACGCTGGCAGGAAATCGTCATCTCCGCGTGCGAACAAAGCGGCAGGA
ACACCGTTCCCCCCGTACTGCCCATCATCGGCTACCGTGAAGCACTCGACAAAATGCCGT
CTGAAAGCACCAAGCTGATTATGAGCATCAACCGCGCCCGCAAACTCGGCGACATACGCC
AACCGTCCGGCGCAATCGTCTTTATGGTCGGGCCCGAAGGCGGCTGGACAGAACAGGAAG
AACAACAGGCATTTGAAGCTGGCTTTCAGGCGGTTACACTCGGCAAACGGATTTTACGCA
CAGAAACCGCCCCACTCGCCGCCCTCGCCGCCATGCAGACGCTTTGGGGCGATTTCGCAT
AAACAGAAATGCCGTCTGAAACCCGTTCAGACGGCATTTTGCAGCCGATTAAGATAGTAG
GTTCAAATAAGATTTCCCGTGTCGTCATTCCCGCGAAAGCGGGAATCTAGAAACGAAAAA
CTACAGAGATTTATCCGAAACAACAACCCTCTCCGCCGTCATTCCCGCAAAAGCGGGAAT
CTAGAAACGAAAAACTACAGGGATTTATCCGAAACAACAAACCCTCTCCGCCGTCATTCC
CGCGCAGGCGGGAATCTAGAAACGAAAAACTACAGGGATTTATCCGAAACAACAAACCCT
CTCCGCCGTCATTCCCGCGCAGGCGGGAATCTAGAAATTTAACGTTGCGGTGATTTATCG
GAAATGACTGAAACTCAACGGACTGGATTCCCGCCTGCGCGGGAATGACGAGATTTTAGG
TTTCTGTTTTTGGTTTTCTGTTCTCGCGGGAATAACGGAATTTTAAGTTTTAGGAATTTG
TCGGAAAAACAGAAATCCCCCCGCCGTCATTCCCGCAAAAGCGGGAATCTAGAAACGAAA
AACTACAGGGATTTATCCGAAACAACAAACCCTCTCCGCCGTCATTCCCGCGAAAGCGGG
AATCTAGAAATTTAACGTTGCGGTGATTTATCGGAAATGACTGAAACTCAACGGACTGGA
TTCCCGCCTGCGCGGGAATGACGAATTTTAGGTTTCTGTTTTTGGTTTTCTGTTCTCGCG
GGAATAACGGAATTTTAAGTTTTAGGAATTTATCGGAAAAACAGAAATCCCCCCGCCGTC
ATTCCCGCGAAAGCGGGAATCTAGAAATTTAACGTTGCGGTGATTTATCGGAAATGACTG
AAACTCAACGGACTGGATTCCCGCCTGCGCGGGAATGACGAATTTTAGGTTGCTGTTTTT
TGGTTTTCTGTTTTTGCGGGAATGACGAATTTTAGGTTTCTGTTTTTGGTTTTCTGTTCT
CGCGGGAATAACGGAATTTTAAGTTTTAGGAATTTGTCGGAAAAACAGAAATCCCCCCAC
CGTCATTCCCGCAAAAGCGGGAATCTAGAAATTTAACGTTGCGGTGATTTATCGGAAATG
ACTGAAACTCAACGGACTGGATTCCCGCCTGCGCGGGAATGACGAAGTGGAAGTTACCCG
AAACTTAAAACAAGCGAAACCGAACGGACTAGATTCCCGCCTGCGCGGGAATGACAGTGT
ATCCATTTCTAATTTTAATCCGCTATATTTTACACAAACTATTTGAACGATATGACCCGC
CTGCCGTAAGCTTTCTCAAGCTCCGCCTGCCTTTGACGCTCCATTCTTTTCTTCTTTTCC
CTACCGAATTTACCCAAAGCACGCTTCAAGTCAAACATCACCTTCAACGAACGGCGGTGT
CTTCTTTCTTGTTCCCTATCTTTTTCCAAATCGCTACCCAACATACTGTTTTTACTGAGG
AACTTGGCATAATGCAATTCTTGGGTACATAAGGCGGGATTAACCTGATAAACAGGCATC
CCCTCCTTATCAAAGAAATAAGTAAACATCATCCAATCTACCGCTTTAATCCACTCTGCC
GGCAAAACGGCAAACCTTTCCAAGAAAAACCGCATCGCCTCACGCGAAATGATATAGCCA
GCCGTCCCCCAATGTTCGCTCTCCAGCAAAGGAAATGACCGATTCTCATAATTCAGGACT
TTATCCGGTCTGACAATAACTTTCGCAAACATCGTTTCCAAACGAACGATAAAGGCAGAA
TCCTTATCAAAACGCTCTTCCAACCAAGTATCTTCGGCAAGGAACTTTTCTGCGTCTTTG
CCAAGCAGGACATCATCCTCAAATACGGCAACATAGGGCAGACCTTCATCCAATGCCTGT
TTCCACAATACGGCGTGGCTCATAAAGCAGGCTTTTTCCACTTCGCTCAACAGGTGCTGT
TTTGCCAATCCCGGCACCAATTCCGCCATCATCCGATTCAGTTCTTCAGACGGCATCAGT
GCGTCGAAAAACTGAAACGGATGCCGCGCACGCCGAAGGTTGCGGCAATGTGCGCCCTG
CGTTCTGCGGCGGAAGCTAAGCTGATAACATGGTTTTGCATAATTTATCCTGTTTTTTGT
CTGTTGGATAAAGCGGCGTTTTTCAACGGTTTTTCAGCAATCGGCGCAAAATGCCGAAGT
ATTGCCTCAAGGTAAACAGCCGCCGCATCCTGCCGTCTGCTGCAAATACGATGTCCATCT
CTCCTCCTTTTATTGGAAAGGCACAATGAACTGTTCGCGCCTTTGCCGGCGTTTTTCCCT
TTCCCTGCTGATTTTGGTCAAGGCGCGGATCAGGCGGTGTTTGAATGTGTTGGCGGGGA

ATCGCGCCTTTGCTGTTTGCGGTTCAGGAGGCGGTCGTGTTCGATCAGGCTGCCCAATGC
GCTGTTTTGGTCGTGAAACTTGGCATAATGCAGCTCTTGGGCGCACAAGGCGGGATTGAG
CTGGCAAACCGGCATTCCTTCCCTGTCGAAAAAATCGCTGAACATCATCAGATCGACGGG
GTGCAGCCCTTCGGGCGGCAGGGCGGCAAACCTGTCCAGGAAAAACCGCATCGCTTTTCG
GGAAATGATATAGCCCGCCGTCCCCCAGTGTTCGCTTTCCAACAGCGGAAAGGCGCGCCC
GCAGTAATCCGCCACGCCGGAGGGCGAGGTCAGGACGTGCATAAACATCGTTTCCAAGCG
GACGATAAAGGCGGTATCCGGGTCAAAGCGTTCTTGCAGCCAAGCGTCTTCGGCAAGGAA
TTTTTCCGCACCTTCGCCGAGTAAAACGTCGTCCTCAAATACGGTGATATACGGCAGACC
TTCGTCCAATGCCTGCTTCCACAATACGGCGTGGCTCATAAAGCAGGCTTTTTCCACTCC
GCTCAAATAGGGGTGCGCCGACAAGCCGGGGACGAGTTCCGCCATTGCCTGTTCCAGCCT
TTCAGACGGCATCAGTGCGTCGAAAAACTGAAACGGGATGCCGTGCCTGCCGAAGGTATC
GGCAATGTGCGCCCTGCGTTCTGCGGCGGAAGCTAAGCTGATAACGTGGTTTTGCATAAT
TTATCCTGTTTTTTGTCTGTTGGATAAAGCGGCGTTTTTCAACGGTTTTTCAGCAATCGG
TGCAAAATGCCGAAGTATTGCCTCAAGGTAAACAGCCGCCGCATCCTGCCGTCTGCCGCA
AAATCCAGCCACGCGCCGGCGGGCAGCGTGTCCGTCCGTTTGAAGCATTGGTACAAAAAC
CGGCGGGCGCGTTCAAAATCTTCTTCCGGCAAATGTTTCTCCAGCAATTCATACGCTACT
GCTTTTATTTGGCGGTATTCAAGGCTGTCGAACCGGGTTTTAAAACCCATAGACTGCAAA
AAATCGTTTCTGGCGGTTTTTTGGATGCCTTGCGCGATTTCGTGTTGGCGGATGCTGTAT
TTGGATGAAACCTGATTGGCGTGAAGGCGGTATTTGACCAAGGCTTCGGGATAATAAGCC
AGCCTGCCCAATTTGCTGACATCGTACCAAAATTGGTAATCTTCCGCCCAATCCCGCTCG
GTGTTGTAACGCAAACCGCCGTCAATGACGCTGCGCCTCATAATCATCGTGTTGTTGTGT
ATGGGGTTGCCGAAAGGGAAAAAGTCGGCAATGTCTTCGTGTCGGGTCGGTTTTTTCCAA
ATTTTGCCGTGTTCGTGGTGCCGCGCCAGCCGGTTGCCGTCCTTTTCTTCCGACAAAACT
TCCAGCCACGCACCCATCGCGATGATGCTGCGGTCTTTTTCCATCTCACCCACGATTTTC
TCAATCCAGTCGGGGGCGGCAATATCGTCTGCATCGGTGCGCGCAATATATTCCCCCCCC
CCCCCCGACTTTGCCAATTCATCCAGCCCGATGTTTAAAGAGGGAATCAGACCGGAATTG
CGCGGCTGCGCGAGGATGCGGATGCGGCCGTCCTGTTCTTGGAAACGCTGGGCAATGGCA
AGCGTACCGTCCGTCGAGCCGTCATCGACAATCAAAATATCCAAGTTGCGCCAAGTTTGA
TTCACGACGGCGGCTAATGATTGGGCGAAATATTTTTCTACGTTGTAGGCGCAAATCAAT
ACGCTGACTAAAGGCTGCAATTTATTCTCCCGATAGGCACGATGCCGTCTGAAGGCTTCA
GACGGCATTTGGACTGTACAACGGTTACTCGCCCAAAAGCGCGATATCCGCTACCGCGTT
CATTTGTTCTGCCAAGCGGTTCAGCAGGTTCAGGCGGTTTTGTTTCACGGCGGCATCTTC
CGCCATCACCATCACGCCGTCGAAGAAGGCATCGACTTGCGGTTTGACGGAAGCCAGTTC
GGACAAGGCGGTCTGGAAATTGCCTTCGGCAACGGCGGCGGCAATTTTCGGCTGCAAGCC
TTGTGCGGCGGCAAAGAGGGCTTTTTCTTCGTCCTGTTGCAGCAAGCTTTCGTTAACCGC
GCCCAACTCGGCATCGGCTTTTTTCAGCAGGTTTTGCACGCGTTTGTTGGCAGCGGCGAG
CGCGGCGGCTTCGGGCAGTTGTTTGAACGCGGCGACAGCCTGCAGTTTGGCGGTCAAATC
GTCCAAACGGCGCGGCTGCTTGGCAAGTACGGCGGCAACGATGTCTTGCGGATAATCGTT
TTGCAGCAATACGGCAAGGCGCGCCTGCATGAAGTCGGCGGTTTCAGACGGCGTTTTTTC
GTTGAGCAAACCTTGCGGGAAGCTGTTGAAGGCCGTCTGAATCAGTTCGTTTACGTCCAA
ACCGTACTGCATCAGCATACGCAAAATACCCAATGCGGCGCGGCGCAGGGCGTATGGGTC
TTTGTCGCCGGTCGGAATCAGGCCGATACCCCAAATGCCGACCAAGGTTTCCAGTTTGTC
GGCAAGCGCAACGGCGGCGGCAATTTTGCCCTCAGGCAGGTTGTCGCCGGCAAAACGCGG
TTGGTAGTGTTGCTCGACGGCTTCGGTAATTTCTTCGGTTTCGCCGTCCAAGCGGGCGTA
GTATTTGCCCATCGTGCCTTGCAGTTCGGGGAACTCGCCGACCATTTCGGTTACTAAGTC
GGCTTTTGCCAAACGCGCGGCGCGTTCGGCTGCGGCGGCATCCGCGCCCAAAGCCTTGGC
GATATGGGCGGCGATGCTTTGCAGGCGTTCGATGCGTTCGGCTTGCGAACCGATTTTGTT
GTGATAAACCACGTTCGTCAGTTTGGGCAGGCGGCTTTCCAAAGTCGCTTTTTGGTCTTG
TTTGTAGAAGAACTCGGCATCAGACAGGCGCGCGCGCAAGACACGTTCATTGCCTTGGAT
GATGTGTGACGGATCTTCGGTTTGCAGATTGGACACCAGCAGGAAGCGGTTCATCAGCTT
GCCGTTTTGGTCGAGCAGCGGGAAGTATTTTTGGTTTTGCTGCATCGTCAGAATCAGGCA
TTCTTGCGGTACGGCGAGGAAGTGTTCTTCAAAACCGGCTTCCAATACCACAGGCCATTC
GACCAGCGCGGTTACTTCGTCCAACAAGGCTTCATCGGCGGCGGCGGTCGCGTTCAGACG
GCGTGCCTGCCCTTCCAATACCGTCTGAATCGCGGCTTTGCGCTCGGCAAACGAAGCGAC
GACTTTGCCTTGCTCGCGCATTTGTGCGGCGTAGCTGTCGGCGTTTTCAATGGTAATTTC
GCCGTCGGAGAGGAAGCGGTGTCCCAAGGTTTTGTTGCCGCTTTGCAGACCCAAAACGCT
GACGTTCACAATGTCGCCGCCGTGCAGTACAACTAGCCCGTGAACGGGGCGCACAAAGGT

AAACGTGCTGCTGCCCCAACGCATTACTTTGGGAATCGGCAGCTTCTTAACCGCTTGATT
GATAATGTCTTCCAAAAGTCCGCCCAACGGTTTGCCGATTTGGACGTATTCGTAGGCGTA
CACGTCCTGCTTGCCGTCGTGGACGATGGTCAAGTCTTCGATTTTCGCGCCCGCACCGCG
TGCGAAACCTTCCAAAGCCTTGGTTGGCGCACCGTCTTTCATGGCATTCGCTACGGCAGG
GCCTTTTTTCACAATTTTTTGATCAGCCTGAACGGCTTTGACGTTTTTGACTTGAACCGC
CAAACGGCGCGGCGAGGCATAAGCCGTAAATTCGGCTGCGCCGTCAACCAGTTGCGCTTT
TTCCAAGCCTTCGGCAACGGAAGCGGCGAAATGGTTGCCCAGATTATTCAGGGCTTTGGG
CGGGAGTTCTTCGGTAAGGAGTTCGATTAAAAGGGTTTGGGTCATCATTCGGCTTTCTTT
GAATTTGGTTAATCTGCCTGTTTATAGGTTTCGCTGTAATTTTCCCAGCCGTCATCCCCA
TAAAAACCGTCAACCAGCGGGGTGGCGTACAAAGTGGCAACATCTTCGCGGTCTGCCAGC
CAAGAGATAATGGCTTTTTTCTCGGTTTCTCCCAAGCTTCGGGCACCGGATTTTTGAAAC
AGGCACGAAAAATCGCCGCAATCGCCCCCCCGCCATTTCAAAGCCGTTTGCCGCAAGATA
CGCAATCAGCTCGTCCATAAAGCGGTCGAACGCTTCGGCATCGTCCTCAGCTTGGTGCAA
ACTGCCTTGAACGCCGAAAATCAATGTTTGAAACTCGCCCAAATGCAGCTTTTTATGCTG
GCGGCGGTTCATTTTGTGCAGGCGTTTCCTGCTTGGGGTGCGGAAATAGACAGGCATGAT
TTTCCTAAAAAATATAATGGCTTCCGGACGGCTGCCTTATCGTGCCGCCCGAACGTAAAA
AATCGTCGCCCCCTTAGGCGGCGTTTGCCTTCATTAAAGGGAAGCCCAGTTTTTCGCGGC
TTTCAACATATTTTTGCGCCACGGCGCGGCTCAATGCACGAATACGTCCAATATAAGTTG
CCCGCTCAGTTACGGAAATCGCGCCGCGTGCGTCTAAAAGGTTGAACGTATGCCCCGCTT
TGAGGACAAGCTCGTAGGCAGGCAGGGCGAGGGCGGCGTTTTCTTCGGCAAGCAGGCGTT
TGGCTTGCGCTTCGTAGTCGTTGAACTGGCGCAGCAGCCAGTCGGCATCGCTGTATTCGA
AGTTGTAGGTGGATTGCTCGACTTCGTTTTGGTGGTACACGTCGCCGTAGGTGACGGTGT
TGCCGTCGAGCGTTTTTGCCCAAACGAGGTCGTAGACGTTTTCTACACCTTGCAAGTACA
TCGCCAAGCGTTCGATGCCGTAGGTGATTTCGCCGAGTACGGGCGTGCAGTCGATGCCGC
CGACTTGTTGGAAATAGGTAAACTGGGTTACTTCCATGCCGTTGAGCCAGACTTCCCAGC
CCAAACCCCACGCGCCGAGGGTGGGGTTTTCCCAGTCGTCTTCGACAAAGCGGATGTCGT
GGACTTTGGGATCGATGCCCAATTCGCGCAGAGAGTCGAGATAGAGGTCTTGGATATTGG
CGGGAGCGGGCTTGAGGGCGACTTGGAATTGGTAATAGTGTTGCAGGCGGTTGGGGTTGT
CGCCGTAGCGGCCGTCTTTGGGGCGGCGGCTGGGTTGGACGTAGGCGGCAAACCAAGGCT
CGGGGCCGAGTGCGCGCAGGCAGGTGGCGGGATGGGATGTGCCGGCACCGACTTCCATGT
CGAAGGGTTGGATGACGGTGCAGCCTTTGTCTGCCCAGAATGTTTGCAGTTTGAAGATGA
TTTGTTGGAAGGTAAGCATGGCTTATGATTCGATAAAATAAAGGGTTTATTTTACTGTTT
CCATTGCTGTTTGGATAGGTTTATCTCAAAGACAGACTGATTTGAAAACACGGCATACAT
GATATAGTGGATTAAATTTAAACCAGTACAGCGTTGCCTCGCCTTAGCTCAAAGAGAACG
ATTCTCTAAGGTGCTCAAGCACCAAGTGAATCGGTTCCGTACTATTTGTACTGTCTGCGG
CTTCGTCGCCTTGTCCTGATTTTTGTTAATCCGCTATATGTTTCGGTTAGGCGGCAGGCT
GCCCTATTGAATACCTTAAAGCAGGCTATGCCTGCCAACGCCATATCCAAACACAGTCTT
TAATTTAAATCCGGAAAATAAAAAGCACGACCAAACGGTCGTGCTTTTCCAAACCAAACA
AGTTTATTTCTTGTGCGAACGGATATAGTCCAAAGTTTTGAGCTGTGCAATCGCAGCAGC
CAATACTTTATGCGCTTCCGCCAAAGCCTTATCGTCTTTAGCTTGGGAAATGCCCGCTTC
TGCGGCTTTTTTCGCCTCTTCCGCACGTGCCCGATCCATCTCCGCACTGCGGACGGCAAC
ATCCGCCAAGACAGTTACTTTATCAGGCTGTACTTCCAAAACACCGCCGGAAACAGCAAC
CAAAACCTCTTTATCCTCGCCCGGAACGGTCAAACGCAAAGCCCCCGGCCGCACCAAACT
CATAATCGGCTCGTGTCGCGGATAAATACCGAGTTCGCCCTGTACAGTCGGAACAACGAT
AAATGTTGCCTCGCCTGAATAGATTTTCTGCTCGCTACTTACCACCTCAACTTGCATGAT
GCTCATGCCGACCTCCTTAGTTTAAGGTTTTCGCTTTCTCTACTGCTTCTTCAATGCTGC
CGACCATATAGAATGCCTGCTCGGGCAGATGATCGTATTCGCCGTTCAAGATGGCTTTGA
AGCCGGCAATGGTATCGCGCAGGGCGACATATTTACCCGGAGAACCTGTAAACACTTCGG
CAACGTGGAACGGTTGGGACAGGAAGCGTTGGATTTTACGCGCACGCATTACGGTCAGTT
TGTCTTCATCAGACAATTCGTCCATACCCAAGATGGCGATGATGTCGCGCAATTCTTTGT
ATTTTTGCAGGGTGGACTGCACACCGCGCGCCACGTCGTAGTGCTCTTGACCCAATACCA
TCGGATCCAGTTGGCGCGAAGTAGAATCAAGCGGATCGACTGCCGGGTAAATACCCAAAG
AGGCAATATCGCGGCTCAATACGACGGTTGCGTCCAAGTGGGCGAACGTTGTTGCCGGAG
ACGGGTCGGTCAAGTCGTCCGCAGGTACATATACGGCTTGGATGGAAGTAATAGAACCGG
TTTGGGTAGAGGTAATACGCTCCTGCAAACGACCCATTTCTTCTGCCAATGTCGGTTGGT
AGCCCACTGCAGACGGCATACGACCCAACAATGCGGATACTTCGGTACCAGCCAGGGTGT
AACGGTAGATGTTGTCCACGAAGAACAATACGTCGCGGCCTTTGCCGTTTTCGTCTTTTT

CGTCACGGAAGTATTCCGCCATGGTCAAACCGGTCAATGCGACGCGCAAACGGTTGCCCG
GAGGTTCGTTCATCTGACCGTAAACCATTGCCACTTTATCCAATACGTTGGAATCTTTCA
TCTCGTGGTAGAAGTCGTTACCTTCGCGGGTACGCTCACCCACGCCTGCGAACACGGACA
AGCCGCTGTGCGCTTTGGCGATGTTGTTGATCAATTCCATCATGTTCACGGTTTTACCCA
CACCGGCACCGCCGAACAGACCTACTTTACCGCCTTTGGCAAACGGACACAGCAAGTCAA
TCACTTTAATGCCCGTTTCGAGCAATTCGGTTGTGGAAGACAGTTCGTCAAACTTAGGGG
CAGCTTGGTGGATGGCACGGCTCTTGTCGGTATCGATCGGACCTGCTTCGTCAACAGGCG
TTCCCAATACATCGACAATGCGTCCCAACGTACCTTTACCTACCGGCACAGTAATGGGCG
CACCGGTATTGCTCACAGTCATGCCGCGTTTCAAACCGTCCGAGCTGCCCATCGCAATGG
CACGGACTACGCCGTCGCCCAAAAGCTGTTGGACTTCCAAAGTCAGACCGTTTTCGTCTA
ATTTCAAAGCGTCGTAAACGCGCGGAATCATGTCGCGTGGAAATTCCACGTCAACAACCG
CACCGATAATTTGTACGATTTTGCCTTGGCTCATTATCGTATCCTAATTTCCGTACAGGA
TTCAGACGGCATCAGACAGCCGCCGCACCTGCTACAATTTCTGACAATTCCGTGGTAATC
GCAGCTTGACGCGATTTGTTATATACCAAACGCAACTCTTTGATGGCATTGCCTGCATTG
TCTGTTGCAGCTTTCATGGCAACCATGCGGGCTGCCTGTTCGGATGCCATATTGTCGCTC
AACGCCTGATAAACCACAGACTCTAAATAGCGGCGAACCAGATATTCCAACACTGCAAGT
GCAGTCGGTTCGTAGCGGTATTCCCAGCTGAACGGTGATTTGGGAGCTGAATCGCCAATC
ACGTTCTCACCGATAGGCAGCAATACTTCCATTCTCGGTTCTTGACGCATGGTATTGACA
AAACCCGAATACACCAGATGGATTCTGTCAATTTCATGTTTCTCATACCGTTGGAAGAGT
TCTGTCAAAGGTCCGAGCAGCATTTCCATTTTTGGGGTATCGCCCAAATTTACGGCACTG
GCAACCACATTCAGACCAATGCTCTGACACGCCATCAGACCTTTACTGCCAAAGCATACG
ATTTCCTCTTCAATACCTTGATTCCGATACTCTTGAACTTGTGCCAAAAACTTTTTCAGC
ACGTTGGCGTTCAAACCGCCACACAAACCCTTATCAGACGTAATCAAAATAAAACCGACA
CGTCTGATTTCCCGATGAGATTCCAGTAACGGAATACCATGATCGGTATTGGTTTGCGCA
AGATGGCTCATCACCATACGCACTTTTTCGGCATACGGACGCGCCAAACGCATCCGTTCC
TGAGTCTTCCGCATTTTAGAGGTTGACACCATCTGCATCGCTTTAGTGATCTTTTGGGTA
TTCTGAACACTGCGGATTTTGGTGAGAATCTCTTTTCCTACTGCCATTTCAGACTCCTTT
CACTTCAAGCCTTATGCCTGATAGGCGTAAGAAGATTTGAAGGATTTCATGGCTGCTTCA
AGCGTTTTCTCGCTCTCGTCGGACATTGCACCTGAAGCATTGACGGCTTCCAAAACTTCC
GGATGTTGGGTACGGACAAAGCTCAAAAATTCAGATTCAAAAGCCAGAGCTTTGGCAACC
GGAACATCAGAATACGAACCGTTGTTGATTGCCCAAAGGGTCAAAGCCATTTCAGCCGTA
TTCAACGTACTGAACTGTTTCTGTTTCATCAGTTCGGTTACGACTTCGCCATGCTCCAAT
TGTTTGCGCGTAGCTTCATCCAAATCGGATGCAAATTGCGAGAACGCCGCCAATTCACGA
TATTGTGCCAACGCCAAACGGATACCGCCACCCAGCTTTTTAATCACTTTGGTTTGTGCA
GCACCGCCTACGCGGGATACGGAAATACCGGCATTGATTGCAGGACGGATACCGGCGTTG
AAGAGGTCGGTTTCCAAGAAAATCTGACCGTCGGTAATCGAAATGACGTTAGTCGGAACG
AAAGCAGATACGTCGCCCGCTTGGGTTTCGATAATCGGCAACGCGGTCAGAGAACCGGTT
TTGCCTTTTACTTCGCCGTTGGTCAATTTCTCCACTTCGTGTTCATTGACACGTGCCGCA
CGTTCCAACAGACGGGAGTGCAGGTAGAACACATCGCCGGGATAGGCTTCGCGGCCGGGC
GGACGGCGCAAAAGCAGGGAAATTTGACGGTAAGCCACAGCCTGTTTGGACAAATCGTCA
TAAACAATCAAGGCATCTTCGCCACGATCGCGGAAGAATTCACCCATCGTACAACCGGAG
TAAGGTGCGATATATTGCAATGCCGCCGCTTCAGATGCAGTTGCAGCAACCACGATGGTA
TGCTCCATCGCGCCATGCTCTTCCAATTTGCGGACCACGTTGGCAATAGAAGATGCTTTT
TGACCGATAGCGACATAGATACAGATAACACCCGTACCTTTTTGGTTGACGATGGCATCC
AATGCTACGGCCGTTTTACCTGTCTGACGGTCGCCAATAATCAACTCACGCTGACCGCGA
CCGACAGGAACCATAGAGTCAATCGCCTTCAGACCGGTTTGCATCGGCTGGTCAACCGAT
TTGCGCGCAATCACGCCCGGTGCGATTTTTTCGATAGGGGCGGTCAAAGTTGTATTAATC
GGGCCTTTGCCGTCGATAGGCCGACCCAATGCATCAACGACGCGTCCGACCAGTTCGCGT
CCGACCGGCACTTCCAAGATACGACCGGTACAGGTAACCGTGTCGCCTTCTTTAATGTGT
TCGTACTCGCCCAACACTACGGCGCCGACGGAGTCGCGCTCCAGGTTCATCGCCAAGCCG
AAAGTGTTACCCGGGAATTCGAGCATCTCACCTTGCATTGCATCTGACAAACCATGGATG
CGAACGATACCGTCAGTTACCGAAATTACCGTACCACAGGTACGCACTTCGGCATTTACA
GACAGATTTTCGATCTTGGCTTTAATCAAATCGCTAATTTCAGCAGGATTAAGCTGCATG
AAAACTCTCCTAATTCGTCATAGTCGTGTACAAGGCACTCAATTTGCCTTGTACAGACAA
ATCCAAAACCTGATCACCCACTTCAACTTTTATGCCGCCAATCAGCTCCGGTTCGATTTC
GACAGAGATTTTCAGCTCGCTGTCGAAACGCTTATTCAGCATTTGCACCAACTCGCCGAC
CTGTTTGTCGGTCAACGGATAGGCACTGTAAATGACGGCAGATTTGATATGGTTGAATGA

TAAGGTCAAGTCTTGATATTGAGCATATACTTCCGGCAATATCGACAAACGTTTCTGCCC
GGCCAAGACGATAACAAAGTTTTTCAACTCCTTGTCTTTCAAACCGACCAAATCGATGAG
GATATCTGCTTTTTCTGAAGCATTCGTTTCAGGACGGTCAATCAATGAAGCCACCTTCCC
TTCCTGAACAACCGCCGCAAGTTTTTCCAGTCCGCCCAACCAAGACTCAATTTGGTTTTT
TTCCTGAGCCAGACCGAACAATGCCTTTGCATAAGGTCTGGCAATCGTTGCGAACTCTGC
CATAAGATTACAGCTCCTGTTTCAGGGTATCGAGCAGTTTTGCGTGTTTGGAAGCATCGA
CTTCGCTGCGCAAAATAGATTCGGCACCTTTGACAGCCAACACGGCAACCTGCTCGCGCA
GGGATTCGCGTGCGCGGAACAATTCCTGCTCCACATCGGCCTTTGCCTGAGCTGCAATGC
GCGCCGCCTCGGAAGAAGCCTGTTCTTTGGCTTCTTCGACAATTTTGGCGGCACGTTTTT
CGGCGTTGGCAACCATTTCGGAAACCTGATTACGCCCTTCTGCCAAGAGTTCTGCAACCT
TTTTTTCAGCCTGCTCAAAATCGCTTTTACCACGCTCGGCGGCAGCCAAGCCTTCGGCGA
CTTTTGCGGCACGCTCATCCAAAGCTTTTGCAATCGGCGGCCACACGAATTTCATGGTAA
ACCATACCAAACCGAAAAAGACGATGATTTGAGCGAATAATGTTGCATTGATATTCACGT
TACTTAACCTTCGTACTGGGGTTAATCAAACAGGCTGCGCCTGTACGGAACGGACGAATC
CGTCCTGATTATGCACCTGCAAACGGGTTAACGAAGGCGAACAGCAGTGCAATGGCGACA
CCAATCAAGAATGCGGCATCAATCAAACCGGCAATCAGGAACAGTTTGGTTTGCAGCGGA
CCGATCAGTTCGGGCTGACGGGCAGAAGACTCCAAATATTTAGAACCGACCATTGCGATA
CCGATAGAGGCACCCAATGCACCCAATGCAACGATCAAACCACATGCGATAGCAATCAAA
CCCATTTTAAACTCCTTAAAGAAACAAAGGTTAAACTACAAAAACAAACTACTTAGGAAA
ATCAGTGCGCATCATGTGCCTGTCCGATATAGACGAACGCCAACGCCATGAAAATAAACG
CCTGCAGGGTAATCACCAAAATATGGAAAATCGCCCATGCCAAACCGGCAATAATGTGGA
ATACAAACAGAATCGGATCCATGACTTCGACGCTGCCGGAAGCCGCCCAAGCACCGCCAA
GCAAGGCTATCAACAAGAATACCAATTCGCCCGCATACATATTGCCGAACAACCGCATAC
CGTGGGATACGGTTTTAGAAAGAAACTCGACCAAATTCAACAGAAAGTTCGCAGGTGCGA
GTTTTGCACCGAACGGCGCGCTGAACAACTCGTGAAACCAGCCACCCAATCCTTTGATTT
TGATGTTGTAATAGATACAAATCAGCAACACGCCGACAGCGAGTGCCAAAGTGGTGTTCA
AATCGGCAGTCGGTACGACGCGCAGCAGGGCGTGATGGTTGCCGGTAATGCCCTGCCATA
CCATCGGCAGCAAATCGACCGGCAGCATATCCATCGCGTTCATCAGAAAAATCCAGACAA
ACAGCGTCAGACCCAACGGCGCGACGGCTTTTCTAGACTTTTCGTTGTGAATGATGCTCT
TACACATATCGTCCACAAACTCAAACAAGATTTCCACTGCGGCCTGGAAACGTCCGGGAA
CGCCTGCCGTCGCTTTTTTTGCACCGCGCCACAACAGAAAGCTGCCGATTACGCCCAACA
GGACGGCAAAAAAGACGGCATCAAGGTTAATAAACGAAAAATCAGCAATGTTTTTCAGTC
CCTGACCCTGAGTAACATCCGACAAACTGGTCAAGCTCTGCAAGTGGTGCTTGATGTAGT
CGGCAGCGGTAATGGTTTCACCTGCCATAATCTTTCACTCTCAACAATACTAAAAAAACC
AAATGGCTGACACCGAGCAGCCCCATCAGAAACGGGGCGAACACCAGCGATTGATGCCAT
ATTGCAAATACGGCAAGCATGGACAACAGCGACAGCACTACTTTTAAAATCTCTCCGAAG
ACGAACATCCTGCTTTGCAGGAAGGGGTTCCCCTGAAAAGTTTTAAAAGTAAAACTGCA
ACAAACGTGGGAAGCAGGTAGGACAAACCGCCACCGACCGCCGAAAGGAATCCGGCAAAA
CCCCATACAGCAAAGGCAACTGCGGCGCATATGGACAATACGGCGGATTGTAGGATGATA
ATCTGCTTCATAAAGGGAATGTTTCCGCCTCGGATTTGGGGCGCGGCTAATATAATTTAG
AAGCCTTATTACGTCAAGCGACAGTTAATCTTTGTGAAACAACGTATCCCAATCCGCCGC
GCTCGCCGCCTGAATAACGGCGACAGGTGTCATTCTAACACACATTACATATAATTACAG
GATATTAAGGAGTTTGTCCGCAATTTCTTTACATTTTTAATGTTCTTACGTGATTTGTTT
GCTTTACGTGGAAATAATAAAAAATCAACGCGAAATTGTAGCAGTTTATCGGTCGGATTG
TCGGCAGTTTGGGGAATTTGCTCAATAAATAAAAGGTCGTCTGAAAATATTTTCAGACGA
CCTTTTCCGAATAAAGGATTAGCAACTGCCTGCCGCTTTAAGCAAAGCATTGCATTGACT
TTTGCCTTTGTGCGTTCCGCCTCCCAAACAAATTGCATCGGAAGTGGTAACGCCGATTGT
GCTGATTACACTGGTAACATAGCATTGGCTCACGCGCTTACCCACAGTTGCGGTAAAGTT
GATGCGTATGCCTTCATTGTTGCGGTTGCTGATTTTTACGGCATTTGGGCTGACGCCCAA
GGCAAACGCGGCACGTTCCTGAAGTTTCTAGTCGGAAACGGTTACATTATTGATTGAGCC
GCAACCTGCTAATGCCAACGCAACGAACGCAGCCGAAACGATGATGCGTGTGTTCATAAT
TTCCTCGAAAATTAAAAATGAAAACAGGAAAACGATTCTTACGTGAAGCAGAAAAAATGT
CAATAGAATTATATTTCCCACTTAAAATCTGGAAAGCTATTCTCTATATTTCAGACGGTA
TATCCCCCAAAATTAAGGCCGGTAATCTATGCCCAACTGCTCCAGCAGGTGGCCGAACGT
TTCAGGCGTATCGAAATACAGGACAATCCTGCCTTTTTTGTGGTTGGCGGTTTTGACTTC
AGCGTTGACACCCAGTTTTTCAGTCAGCAAATCATTCAGGCGGCCGATGTCGGCGGCGGC
AGTCTTTTTGGGCTCGGGACGTTTGTTTTGAAGGGCGGCCTGGCTGCGGCGTTCGACTTC

GCGCACCGACCAGCCGTTTTTGACGGCCTTTTGCGCCAATTCGAGCTGTTCGACGAGTTT
TCGGAGAGGACCATGCGCGGGCGTGCCCCATTTC

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 5>:

gnm_5

CAGACATTACCGTGTACAACGGCCAACACAAACGAAGCAGCACAAGCCGTTGCAGATGCC
TTTACCCGGGCTACCGGCATCAAAGTCACACTCAACAGCGCCACACGCGACCAGCTTGCC
GGTCAAATCAAAGAAGAAGGCAGCCGAAGCCCCGCCGACGTATTCTATTCCGAACAAATC
CCGGCACTCGCCACCCTTTCCGCCGCCAACCTCCTAGAGCCCCTGCCCGCCTCCACCATC
AACGAAACACGCGGCAAGGGCGTGCCGGTTGCCGCCAAAAAAGACTGGGTGGCACTGAGC
GGACGTTCGCGCGTCGTCGTTTACGACACCCGCAAACTGTCTGAAAAAGATTTGGAAAAA
TCCGTCCTGAATTACGCCACGCCGAAATGGAAAAACCGCATCGGTTACGCCCCCACTTCC
GGCGCGTTCTTGGAACAGGTTGTCGCCATCGTCAAACTGAAAGGCGAAGCGGCCGCATTG
AAATGGCTCAAAGGTCTGAAAGAATACGGCAAGCCTTACGCTAAAAACTCCGTCGCCCTT
CAAGCGGTTGAAAACGGCGAAATCGATGCCGCCCTCATCAACAACTACTACTGGCACGCT
TTTGCGCGTGAAAAAGGCGTACAAAATGTCCACACCCGCCTGAATTTCGTCCGCCACAGA
GATCCCGGCGCACTCGTTACCTATTCCGGCGCAGCCGTGTTAAAATCCTCCCAAAACAAG
GATGAGGCGAAAAAATTCGTCGCCTTCCTCGCCAGCAAGGAAGGACAGCGCGCCCTGACC
GCCGTCCGTGCCGAATATCCTTTGAATCCGCACGTGGTATCCACTTTCAATTTGGAACCC
ATCGCCAAGTTGGAAGCACCCCAAGTGTCCGCCACCACTGTTTCCGAAAAArAACACGCC
ACCCGGCTGCTTGAGCAAGCCGGTATGAAATAAGCCGTTTTCGGATTGTCAAACGGGTGG
ACATTTATACTGTCCGCCCGTTTTGCCGATAAAAAACACTATGTCTCCTAAAAAAATACC
CATTTGGCTTACCGGCCTCATCCTACTGATTGCCCTACCGCTTACCCTGCCTTTTTTATA
TGTCGCTATGCGTTCGTGGCAGGTCGGCATCAACCGCGCCGTCGAACTGTTGTTCCGCCC
GCGTATGTGGGATTTGCTCTCCAACACCTTGACGATGATGGCGGGCGTTACCCTGATTTC
CATTGTTTTGGGCATTGCCTGCGCCCTTTTGTTCCAACGTTACCGCTTCTTCGGCAAAAC
CTTTTTTCAGACGGCAATCACCCTGCCTTTGTGCATCCCCGCATTTGTCAGCTGTTTCAC
CTGGATCAGCCTGACCTTCCGTGTCGAAGGCTTTTGGGGGACAGTGATGATTATGAGCCT
GTCCTCGTTCCCGCTCGCCTACCTGCCCGTCGAGGCGGCACTCAAACGCATCAGCCTGTC
TTACGAAGAAGTCAGCCTGTCCTTGGGCAAAAGCCGCCTGCAAACCTTTTTTTCCGCCAT
CCTCCCCCAGCTCAAACCCGCCATCGGCAGCAGCGTGTTACTGATTGCCCTGCATATGCT
GGTCGAATTTGGCGCGGTATCCATTTTGAACTACCCCACTTTTACCACCGCCATTTTCCA
AGAATACGAAATGTCCTACAACAACAATACCGCCGCCCTGCTTTCCGCTGTTTTAATGGC
GGTGTGCGGCATCGTCGTATTTGGAGAAAGCATATTTCGCGGCAAAGCCAAGATTTACCA
CAGCGGCAAAGGCGTTGCCCGTCCTTATCCCGTCAAAACCCTCAAACTGCCCGGTCAGAT
TGGCGCGATTGTTTTTTTAAGCAGCTTGTTGACTTTGGGCATTATTATCCCCTTTGGCGT
ATTGATACATTGGATGATGGTCGGCACTTCCGGCACATTCGCGCTCGTATCCGTATTTGA
TGCCTTTATCCGTTCCTTAAGCGTATCGGCTTTAGGTGCGATTTTGACTATATTATGTGC
CTTGCCCCTTGTTTGGGCATCGGTTCGCTATCGCAATTTTTTAACCGTTTGGATAGACAG
GCTGCCGTTTTTACTGCACGCCGTCCCCGGTTTGGTTATCGCCCTATCCTTGGTTTATTT
CAGCATCAACTACACCCCTGCCGTTTACCAAACCTTTATCGTCGTCATCCTTGCCTATTT
CATGCTTTACCTGCCGATGGCGCAAACCACCCTGAGGACTTCCTTGGAACAACTCCCAAA
AGGGATGGAACAGGTCGGCGCAACATTGGGGCGCGGACACTTCTTTATTTTCAGGACGTT
GGTACTGCCGTCCATCCTGCCCGGCATTACCGCCGCATTCGCACTCGTCTTCCTCAAACT
GATGAAAGAGCTGACCGCCACCCTGCTGCTGACCACCGACGATGTCCACACACTCTCCAC
CGCCGTTTGGGAATACACATCGGACGCACAATACGCCGCCGCCACCCCTTACGCGCTGAT
GCTGGTATTATTTTCCGGCATCCCCGTATTCCTGCTGAAGAAATACGCCTTCAAATAACA
GCTTGAGGAAGTACCGCCATGACCGCCGCCCTGCACATCGGACACCTGTCCAAAAGTTTT
CAAAACACCCCAGTTTTAAACGACATTTCGCTCAGCCTCGACCCGGGCGAAATCCTCTTT
ATCGTCGGCGCGTCCGGCTGCGGCAAAACCACCCTTTTACGCTGCCTTGCCGGTTTTGAA
CAACCCGATTTTGGCGAAATTTCGCTTTCCGGCAGAACCATCTTCTCGAAAAATACCAAC
CTCCCCGTCCGCGAACGCCGTTTGGGTTATGTCGTACAGGAAGGTGTGCTGTTCCCCCAC

CTGACCGTTTACCGCAACACCGCCTACGGGCTGGGCAACGGCAAAGGCAAGACGGCGCAA
GAGCGGCAGCGCATCGAAGCTATGTTGGAATTGACCGGCATTTCCGAACTTGCCGGACGC
TATCCGCACGAACTTTCGGGCGGACAGCAACAGCGCGTCGCCCTCGCCCGCGCCCTCGCA
CCCGATCCCGAACTGATTTTGTTGGACGAACCCTTCAGCGCGCTGGACGAACAGTTGCGC
CGCCAGATTCGCGAAGACATGATTGCCGCCCTGCGCGCCAACGGCAAATCTGCCGTTTTC
GTCAGCCACGACCGCGAAGAAGCCCTGCAATACGCCGACCGGATTGCCGTGATGAAACAG
GGGCGCATCCTCCAAACTGCAAGCCCTCACGAATTGTACCGACAACCTGCCGACCTTGAT
GCCGCCCTGTTTATCGGCGAAGGCATCGTGTTCCCCGCCGCGCTCAACGCCGACGGCACC
GCCGATTGCGGATTGGGCCGCCTGCCCGTTCAAAGCGGCGCACCCGCAGGCACACGCGGC
ACACTGCTCATCCGTCCGGAACAGTTCAGCCTTCACCCCCATTCCGCACCCACCGCCTCC
ATTCACGCCGTGGTTCTCAAAACCACGCCCAAAGCGCGGCATACCGAAATCAGCCTCCGG
GTCGGACAAACCGTCCTCACGCTCAACCTCCCTTCCGCCCCCACCCTGTCAGACGGCATT
TCCGCCGTCCTCCATTTGGACGGTCCCGCCCTGTTCTTCCCCGGAAATACCCTCTGAAAG
GCGGCAGCATCCACAAGCCTGCGGATATTTATCTTGTTGGAAACAGAATTTGTTTGCTAT
ATTCAACCTGCGCGCCTCAAGCCAAACACGGCACGCACGGCACGCAGGCAGCCGTTTCTG
CCTATATGCCGCCCCTTCCAACCACATATGCCGCACACCGCAGCATACGAAAGGATATAT
CATGGCAAAAGTACTCATCGTACCCGTATCTGCCGGACTGGACGCCTCCGCCGCCGCACA
AGCCTTTGCAAAAGCACTGGACGCACAAATTTTCCAAGCCGTTGACGCAACCGCCGAAAC
CCTGCTCGCGCAAGGCAAAAGCGACGACTGGTTCGACGCACTGGTCGGCAAAGTTGCCGC
ACTCGATGCCGCCAACCTCGTCATCGAAGGCATCGCGCCCGATGCCGACAAAATCTACCT
CGCAGGCAAAAACGTCGAACTGGCATTGTCCCTTGACGCGGCAGCCGTCTTCGCCGTCCG
TTCCGACAACGCCGATGCCGACGAACTGGCAAATCGGGTGAACCTTGCCAAACAGTTCTT
CGCCGCCGCGCCGGGCGTATTGGAAGGTTTTGTCGTGGACGGCGCGGCAGCCTCCGTTGC
CGAAGCGGCAGCCGAAAAAACCGGCCTGACCTTCTTCGGTTCGAGCGACGCGCTGAAAGA
CGTATCCGTATTGGCAGGCCGCGAAGCAAAACGCCTGTCGCCGGCGCAATTCCGCTACAA
CCTGATCGACTTCGCCCGCCAAGCCGACAAACGCATCGTCCTGCCTGAGGGCGCAGAACC
CCGCACCGTCCAAGCCGCCGCCATCTGCCACGAAAAAGGCATTGCCCGCTGCGTCCTGCT
TGCCAAACGCGAAGAAGTCGAAGCCGTTGCCAAAGAACGCGGCATCAGCCTGCCCGACTC
TTTGGAAATCATCGATCCCGCCTCATTGGTCGAACAATACGTCGAGCCGATGTGCGAACT
GCGCAAATCCAAAGGCCTGACACCCGAAGACGCGCGCAAGCAACTGCAAGACACCGTGGT
ACTCGGTACGATGATGATGGCGCAAAATGATGTGGACGGTTTGGTATCCGGTGCGGTTCA
CACCACCGCCAACACCATCCGCCCCGCTTTGCAACTGATTAAAACCGCACCGGGCGCAAG
CCTCGTGTCCAGCGTATTCTTTATGCTGCTGCCCAACCAAGTCCTCGTCTTCGGCGACTG
CGCGGTTAATCCGAACCCGACCGCGCAACAGCTTGCCGACATCGCCATCCAGTCTGCCGA
TTCCGCAAAAGCCTTCGGCATCGACCCGAAAGTGGCGATGATTTCCTACTCCACCGTCAA
CTCCGGCAGCGGCCCCGATGTCGATACCGTCATCGAAGCAACCAAACTTGCCCGGGAAAA
ACGCCCCGACCTCGCCATCGACGGCCCGCTGCAATATGATGCGGCAACCGTGCCGGGTGT
GGGCAAATCCAAAGCTCCGGGCAGCCCGGTGGCAGGACAGGCAACCGTTTTGGTCTTCCC
CGACCTGAACACCGGCAACTGCACCTATAAAGCCGTCCAACGCAACGCCAACGTCTTAAG
CGTCGGCCCGCTGCTGCAAGGCCTGCGTAAACCGGTCAACGACCTCTCCCGCGGCGCACT
GGTAGAAGATATCGTGTTTACCATCGCCCTGACTGCCGTTCAGGCAAAACAAATGGAAGG
CTGACAAACGGCTTTCCGGGTTTAAACCCTATGCCGTCTGAAGACAGACTCCCGTTTTCA
GACGGCATTTTTATCAGCACGGCACATTTGTTTGTTAAAATCGCAGCCATATTGCAAAAA
AAGAGGAGGAAGCCATGCAAACCGCCATTATCGATTACGGTATGGGCAACCTGCATTCCG
TATTGAAATCCGTCCGGACGGCGGGGCAGCTTGCCGGAAAAAATACCGAAATCTTTTTAA
GCGGCGACCCCGACCGCGTGTCCCGCGCCGACAAAGTCATTTTTCCCGGTCAGGGCGCGA
TGCCCGACTGTATGGCGGCATTAAAACGAGACGGTTTGGACGAGGCAGTCAAAGATGCCT
TAAAAAACAAACCGTTTTTCGGAATCTGCGTCGGCGCGCAACTTTTATTCGACCACAGTG
AAGAAGGAAACACCGACGGCTTGGGCTGGTTCGGCGGCAAAGTCAGACGCTTTGAGCGCG
ACCTCCGCGACCCGCAGGGATGCCGTCTGAAAGTCCCGCATATGGGCTGGAACACCGTGC
GCCAAACCCAAAACCACCCGCTGTTTAAAGATATTCCCCAAGACACGCGTTTTTACTTCG
TCCACAGCTACTATTTCGCCCCCGAAAATCCCGAAACCATATTGGGCGAAAGCGACTACC
CGTCCCCGTTTGCCTGCATCGTCGGCAAAGACAACGTATTCGCCACGCAATTTCACACCG
AAAAAAGCCACGATGCCGGGCTGACGATGTTGAAAAACTTTTTAAACTGGTAAGCCGGAC
ACGGCCCCGCACAAGGAGAAAAATTATGCTGCTGATACCCGCCATCGATTTGAAAGAAGG
ACGCTGCGTCCGCCTGAAACAAGGGCTGATGGAAGAGGCGACCGTCTTTTCCGATTCGCC
CGCCGAAACCGCGCTGCACTGGTTCAAACAAGGCGCGCGCCGCCTGCATCTGGTAGATTT

GAACGGCGCGTTTGCCGGCGTTCCGCAAAACCTGCCCGCCATCAAAGACATCCTTGCCGC
TGTCGCCAAAGACATCCCCGTACAGCTCGGCGGCGGCATACGCGATTTGAAAACCATCGG
ACAATATTTGGATTTGGGCTTAAACGACGTGATTATCGGCACGGCGGCGGTCAAAAACCC
CGACTTCGTGCGCGAGGCGTGCAAAGCCTTCCCCGGCAGGATTATTGTCGGGCTGGATGC
CAAAGACGGTATGGCCGCCATCGACGGCTGGGCAACCGTAACCGGGCATCATGTAATTGA
TTTGGCAAAACGCTTTGAAGACGACGGCGTCAACAGCATCATCTACACCGACATCGGGCG
CGACGGTATGATGAGCGGCGTGAACATCGACGCGACGGTCAAACTCGCCCAAACCGTCCG
CATTCCCGTCATCTCCTCCGGCGGACTGACCGGCTTGGACGACATCCGCGCCCTGTGTGC
CGCCGAAAAACATGGCGTAGCAGGCGCGATTACCGGCCGCGCGATTTACGAGGGTAGCAT
CGATTTTGCCCAAGCGCAGCAACTGGCAGATTCCCTCGACTAAAGGCATCCGATTATGGC
ACTGGCAAAACGCATCATCCCCTGTCTCGACGTAAAAGACGGGCGCGTCGTCAAAGGCGT
GAACTTCATCGGTTTGCGCGACGCGGGCGACCCCGTCGAAGCCGCCAAACGCTACAACGG
CGAAGGCGCGGACGAATTGACCTTCCTCGACATCACCGCCTCATCCGACAACCGCGACAC
CATCCTGCACATCATCGAAGAGGTTGCCGGACAAGTCTTCATCCCCCTGACCGTCGGCGG
CGGCGTACGCACCGTTGCCGACATCCGCCGCCTGCTCAATGCCGGCGCGGACAAAGTCAG
CATCAACACCGCCGCCGTTACCCGTCCCGATTTAATTGACGAAGCCGCCGGATTTTTCGG
TTCGCAAGCCATCGTCGCCGCCGTCGATGCCAAAGCCGCCAACCCCGAAAACACACGCTG
GGAAATCTTTACCCACGGCGGGCGAAATCCGACCGGTTTGGATGCGGTGGAATGGGCGGT
CGAAATGCAAAAACGCGGCGCGGGCGAAATCCTGCTCACCGGTATGGACAGGGACGGTAC
GAAACAGGGTTTCAACCTGCCGCTGACCCGCGCCGTTGCCGAAGCCGTCGACATCCCCGT
CATCGCCTCCGGCGGGGTCGGCAATGTCCGGCACCTGATTGAAGGCATAACCGAAGGCAA
AGCCGATGCCGTACTTGCCGCCGGCATTTTCCATTTCGGGGAAATCGCCATCCGCGAAGC
CAAACGCGCTATGCGCGAAGCCGGCATCGAAGTGCGCCTCTGACCGCCTCGACTATGCCG
TCTGAAAGGAAATATGGATAAAAACCTGCTTGAAGCCGTCAAATTTGACGAAAAAGGTTT
GGTTTGCGCCATCGCCCAAGATGCCGAAACCAAACGTATTTTAATGGTGGCGTGGATGAA
CGCCGAAGCCCTGCAAAAAACCGTCGAAACCGGCTTTGCCCACTATTACAGCCGTTCGCG
CCAAAAACAATGGATGAAGGGCGAAGAGTCGGGACACACGCAAAAAGTCCGCGCACTGCG
CCTCGACTGCGACGGCGACGCCATTGTGATGCTCATCGCCCAAAACGGCGGCATCGCCTG
CCACACCGGGCGAGAAAGCTGCTTTTACAAAGTCTGGCGTGGCAGCGCGTGGGAAACCGC
CGATGCCGTCCTGAAAGACGAAAAAGAGATTTACGGCAGCACGCACTGACCGCCTCCAAC
ATTGAATTATCAGGCATTTTTTTGTACAATTTCGCCGTCTCAAACACTGTCCGGGCCGTC
TGAAAAGCGGCCTGAACCTTTTTGCAAAGAAAACCATGTCCCAAGAAATCCTCGACCAAG
TGCGCCGCCGCCGCACGTTTGCCATCATCTCCCACCCTGACGCAGGTAAAACCACGTTGA
CTGAAAAACTCTTGCTGTTTTCGGGCGCGATTCAGAGCGCGGGTACGGTAAAAGGCAAGA
AAACCGGCAAATTCGCCACTTCCGACTGGATGGAAATCGAGAAGCAGCGCGGCATTTCCG
TGGCATCAAGTGTGATGCAGTTCGATTACAAAGACCACACCGTCAACCTCTTGGACACGC
CGGGACACCAAGACTTCTCCGAAGACACCTACCGCGTTTTAACCGCCGTGGACAGCGCAT
TAATGGTCATCGACGCGGCAAAAGGCGTGGAAGCGCAAACCATCAAGCTCTTAAACGTCT
GCCGCCTGCGCGATACACCGATTGTTACGTTTATGAACAAATACGACCGCGAAGTGCGCG
ATTCCCTGGAACTTTTGGACGAAGTGGAAAACATTTTAAAAATCCGCTGCGCGCCCGTTA
CCTGGCCGATCGGTATGGGCAAAAACTTCAAGGGCGTGTACCACATCCTGAACGATGAAA
TTTATCTCTTTGAAGCTGGCGGCGAACGCCTGCCGCACGAGTTCGACATCATCAAAGGCA
TCGATAATCCTGAATTGGAACAACGCTTTCCGTTGGAAATCCAGCAGTTGCGCGACGAAA
TCGAATTGGTGCAGGCGGCTTCCAACGAGTTTAATCTCGACGAATTCCTCGCCGGCGAAC
TCACGCCCGTATTCTTCGGCTCTGCGATTAACAACTTCGGTATTCAGGAAATCCTCAATT
CATTGATTGACTGGGCGCCCGCGCCGAAACCGCGCGACGCGACCGTACGTATGGTCGAGC
CGGACGAGCCGAAGTTTTCCGGATTTATCTTCAAAATCCAAGCCAATATGGACCCGAAAC
ACCGCGACCGTATTGCCTTCTTGCGCGTCTGCTCCGGCAAATTCGAGCGCGGCATGAAGA
TGAAACACCTGCGTATCAACCGCGAAATCGCCGCCTCCAGCGTGGTTACCTTCATGTCGC
ACGACCGCGAGCTGGTTGAAGAAGCCTACGCCGGCGACATTATCGGCATCCCGAACCACG
GCAACATCCAAATCGGCGACAGCTTCTCCGAAGGCGAACAACTGGCGTTCACCGGCATCC
CATTCTTCGCACCCGAACTGTTCCGCAGCGTACGCATCAAAAACCCGCTGAAAATCAAAC
AACTGCAAAAAGGCTTGCAACAGCTCGGCGAAGAAGGCGCGGTGCAGGTGTTCAAACCGA
TGAGCGGCGCGGATTTGATTTTGGGCGCGGTCGGCGTGTTGCAGTTTGAAGTCGTTACCT
CGCGCCTCGCCAACGAATACGGCGTAGAAGCCGTGTTCGACAGCGCATCCATCTGGTCGG
CGCGCTGGGTATCGTGCGACGACAAGAAAAAACTGGCTGAATTTGAAAAAGCCAACGCGG
GCAACCTCGCCATCGACGCAGGCGGCAACCTCGCCTACCTCGCCCCCAACCGCGTGAATT

TGGGACTCACGCAAGAACGTTGGCCGGACATCGTGTTCCACGAAACACGCGAACATTCGG
TCAAACTGTAAAAAGCAATCGGCGATAAAATGCCGTCTGAACCCGAAAAAAGGCTTTCAG
ACGGCATTTTTGCCTGCAACTCAAATGCACCGATCAAAATCAAGACGCATCGGATACCGT
TATCGGGCATCCCGTCCTCATGAATTTAGGGCTGACGCAAGAACGCTGGCCGGACATCGT
GTTCCACGAAACGCGCGAACATTCGGTCAAACTTTAAAAAACAATCGGCAATAAAATGCC
GTCTGAACCCGAAAAAAGGCTTTCAGACGGCATTTTTGCCTGCAATTCAAACGCAGACGG
TCAAAATCAAGGCGCATCGGATACCGTTATCGGATGCGTCCCATCCGCATGAATTTGGGG
CTGACGCAAGAACGCCGATGTGATTTCACATCCCGTACTGTTTCGACAGCTTCACATAAT
GCGCGGCGGAATATTTCAAAAAGGCTTTTTCCTCATCGGTCAGCACGCGCACCTGTCTGA
CCGGCGAACCGACATAAAGATAGCCGCCCGCCAAGCGTTTGCGCGGCGGAACGAGGCTGC
CCGCGCCGATCATCACTTCGTCCTCAATCACGGCATCGTCCAGAACCGTCGTCCCCATGC
CGACCAGGACGCGGTTGCCGATACGGCAGCCGTGCAGCATCACTTTGTGCCCCACGGTAA
CGTCTTCGCCGATAACCAGCGGCGATCCTTCGGGTTTGGCGGCGGTTTTGTGGGAAACGT
GCAAGACGCTGCCGTCCTGTATATTGCTGCGCGCGCCGACGGTGATGCTGTTCACATCGC
CGCGCAACACGGCGCACGGCCACACGGAAACATCTTCGGCAAGCGACACTTCGCCAATGA
CGACGCAGGCTTCGTCTATCATACAGGTTTCGTGGATTTCGGGCGTGCGGTTTTGGAAAG
TTCGGATTGCGTTCATTTTTCCTCCTTCGGTAAGGTATATATTGTTAAAGGATTTATTAA
ATATTCCCCCTGATTGCTTTTAAAATCCTGCCTGTAATATCGACCCCGAGTAATGTGATT
ATCGGAATATCAGCTTATATATCAATTTATTGGACTTTAACAGCATAAACCTTAAATGA
TACGCCCTTCTTTTTATATCAGCATCACACTCTATATTTTTACTCGTCATTATAAAAAGC
AAAACGAGATATTCGTAGGAAAGAAAAGAATAAAGATAACTCGATATATCCCTATTAAAT
TCCATTTCCGCATTTTTCTCCAAAATATATAATAATGACTTTATACTTTTTTCCGAAACA
GTTCCCGTAATAGAATCTTTTCTTCCCTGCCGATAATAGTAATAACAACCGTCCAAATAA
GAAAAAGTTGTTGCCGCATTAAATAACCTCATTGACCATTCGATATCTTCAGAATAAATT
CCCCTTTCAAAAAACAGTTTTTCTCTAATAATCAATTCTCTTTTTATAATCTTATTCCAC
GCCGAACCCGGAAATTTTCTAAATCGGCATAATCCTTTCAAAACTTCGACTTTGGATTGA
TTGAGTATTTTTTCAGGCTGATAATCTTCGCCAAAATATGAAACACTTCCCTTATCATAT
TTAACCGCATTTAAAAACACCACATCCGGCATATCAGTATCATCTTTACTAAGAAAATCC
AGCAAAATCTGACAGTTAATAAAATCATCCGAATCAATAAAGACTATATATTTTCCGTTT
GAATTTTTTATTCCGGTATTTCTCGCCTCCGACAACCCTTGGTTATCCTGATATATATAT
TTGATATTTGGTGTCCGGTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGT
TGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTT
TGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTT
GTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTTGTTGTTT
GTTGTTTGTTGTTTGTTGTTTGCTGTTTGATATTTTATCTATATATTTGTAACATATATC
TTCACTTCCGTCTTTTGACCCGTCATTCACAAGGATAAGTTCGACATTTTCATTACTTAA
TATAGATTCTATGGAACTTAAACACGCTTCCAAATAACTTTCGACATTATAAATAGGAAC
TACGATACTTAAATCCATATCTTTCTCATTTTACTAAATCATTTTAATCTTAACCCAATC
ATAACCGGCAGGAGGGGAAACGCCCCCCTGTTTGATAACGGACGCGCCGTTTCCTGCCGC
CCGAAAGGTTTCAGACGGCAGGATTCCGGTTATTTGCCCGCTTTGAGCCCTTGCCACAG
CTTCACGCCCAGTTTGACGGATTCTGCGGATTTGGGCGATACGATGAAACTTTTTTCCAT
CAGTTCTTTGTTCGGGAAAATCGATGCGTCGGAGGTGTATTTTTCATCCATCAGCTCGCG
CGCGGGACGGCTGGCGGGCGCGTAGGTAACGAAGCTGCCGTTTTTCGCCGCCACCTCGGG
CCGGAGCGTGTAGTCGATATAGCGGTGGGCATTGGCAACGTTTTGCGCGTCGCGCGGAAT
CATAAAGGAATCCACCCACACGCCCACGCCGGTTTTCGGGGTCAATACTTTGATTTCCAC
GCCGTTTGCGGCTTCTTCGGCACGGGTTTTGGCAATGTTCAAATCGCCGCCGTAACCGAT
GGCGGCACACAGGTTGCCCGCCGCCATATCGTCGATATAGCCGGAAGAGCTGAAGCGTTT
CACGTCGCCCCGGACGGCTTTCATCATATCGACGGCGGCTTTGATGTCTTCGGGATTCTC
ACTGTTGGGGTCTTTGCCCAAATAGTGCAACGCCAAGGGAATCTGTTCGATTGCGCTGTC
GAAATAGCTGATGCCGCAGGATTTGAGTTTGGCGGTGTATTCGGGTTTGAACACCAAATC
CCATTCGTTTTCGGGCAGCTTGTCCGTACCCAATGCTTTTTTCACCTGCTGGGTATTGAT
TGCCAAGGTATTGATGCCCCAGAAATAGGGGACGGCGTATTCGTTGCCCGGATCGACGGC
TTCCATCATTTCAGCAAATCTTTATCGATGTTGCCGTAATGGGGGATTTGCGCCTTGTC
GATTTTCTGATACGCGCCCGCTTTGATTTGCCGGCCGACGTTGGCGATGGACGGCGCGGT
CAGGTCGTAGCCGGATTTGCCGGTCAGGACTTTTGCCTCCAGTGTTTCGTTGCTGTCGTA
ATAATCGGAACGCGTCTTGATGCCGGTTTCTTTTTCAAAGGCGGCAACGGTTTCGGGATC
GACATAATCCGACCAGTTGTAGATGTTGAGTTTGCCCGATTGTTCGGCTTCGGGCTTGGC

GGAGGGGGTTTGGGCGGCGGTATCGCTTCCGCCGCCGCACGCAGTCAAGGCGAGGCTCAG
GATTGCCGCCGCCACCAGTGTTTTTTTCATGGTTTTCGACCTTTCGTTCAAAAATAAAAA
ACATTGCAGGGACGGACGGGCCGCCTGCATCGGAAGCTGTGGGGGCGGCCGGTCGGCTTT
CTTGGCCCTTGGTGCAGCAAGGTTTGCATTAAACGGCAAAAACAGCCGGGGCGCAACCGT
TAATTTTCACCGGTTTGCCGTTCCGGTGCCCGATCGGGCGGTAATGCGGGACATTGTTCC
GCCTGCGTAAAAATGCCGTCTGAAGCCTCGGCGGCTTTCAGACGGCATCGGGGCGGACGG
CATCAGTTGCTCAACACGTCCACGCCTTTGGGCGGGGTAAACTTGAACGCGCCGCGCGAG
AGTTGGGGATTGGTATTCAAACCGCCGAAACTGATGGAGGTTTGGTTGCCGAAGCTGTCT
TTAAGCTGCATGGCGGCGAGGTTGCCGCCTTTGAAGCCGATGCGGATGTATTGGTAGCCG
GCGTTGTTGCGTTTGGGCGTTGCCAGCACATAATCGATGCCGTTGGACGAACCGTCCTCT
TTCAGCGTGTAGCTGCTTTCGAGGGCGGTTTTGTTCGACAGGATGGCGGCGGGGCTGCCG
CCTATGGCCTGGTCTTGGGACGACTTGGTCACTTGTGCCAGATCAACATCGTAGAGCCAA
ACGGTTTGACCGTCGCCGACGATGGTTTGCCTGTAAGGTTTGGTGTATTCCCATTTGAAA
AGGCCCGGTCGCAGGATTTTGAACGTGCCGTGCGCGGTTTGGGTTTTCTTTTTGCTTTGG
ACGGTTTGGGTGAAGCTGCCGCTGATACCGTCGGCATCGTTGTTGAATTGCTTAAGCGCG
TCTACCGCGCCCGCCTGTGCGGAAGCGACGGCGACGGTCAGGGAGCAAACGGCGAGGAAT
TGGAACAGGTTGTGCGGTTTCATCATTATTTTTCCTTGTCGGGATGAGTGTGGCGTAAAG
TATCGGCGCAGCAAAACAATCATACGGGCGGCGTTACGGGCGGTTTGCATTTTGCAAACC
GCGTTTTCCGAGGGCTGATTTTTTGCCCACCGGAAAAGGCGGCGCGCCACGCTGCCCTTT
AATGTGTGCCGCGTTATAGTGGATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGAC
AGTACAAATAGTACGGAGCCGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCT
TTGAGCTAAGGCGAGGCAACACCGTACTGGTTTTTGTTAATCCACTATATTTGACGGTTT
GGGCGCGGAATGCCGAATCGCGGAGATTATATGCCTCATCTTGTTTTTTTTAAACGCTAT
AATATTTGTTTTCCGAACACGGCGGACTTGAGATGAAACCCGATATTTATGCTTTGCTGG
AACGCGCCCTGCTTTCGGGCGACCCAGATGAAAAAGGACGGCTGACGGATGAGGCGTTTG
CCGCCGTTCAAAATGCGGACGGGGCGGAAACAAACGCACCGCCGGCGGACTTCCCCCGCG
CGGGACGACCGGACAAGCCTGTTTTGGTCGCGCCGTCGCAGCTGACGCCACGCAAAATGA
ACACAACCGAAGGCTATGCGGCGATGCTGCACGCGATTGCGCATATCGAATTCAACGCCA
TCAATCTGGCTTTGGACGCGGCATACCGTTTCCGCACGCTGCCGTTTCAGTTTGTCCGCG
ACTGGGTGAAAGTGGCGAAGGAAGAGGTGTACCATTTCCGCCTGATGCGCGAAAGGCTGC
GCGCTTTCGGCTTCGATTACGGCGATTTTGAAGCACACAATCATTTATGGGATATGGCAT
ACAAAACCGCCTACGATCCTTTGTTGCGTATGGCTTTAGTGCCGCGCGTTTTGGAAGCGC
GCGGGCTGGACGTTACGCCCGGCATACGCGCGAAGGTGGCGCAGCGCGGTGATTCGGAAA
CCTGCGGCGTGTTGGACATCATTTACCGCGACGAAGTGGGACACGTCGCCATCGGCAACC
GGTGGTATCAACACCTTTGCCGCGAACGCGGTTTGGAGCCTGTCGCCCTGTTCCGCAGCC
TGATTGCCCGTTACGATATGTTTATCTTCCGGGGCTATGTGAACATCGAAGCGCGCGAAA
AAGCAGGCTTCAGCCGCTTTGAATTGGATATGTTGGAAGATTTCGAGCAGGGTTTGAAAC
AAAATAAACATGCCGTCTGAAACCCTTCGTCCCGCACTTTATAAAAAAGGAACACACATG
ATACAAGCCGTATTGTTCGACCTCGATGGCACGCTCGCCGACACCGCCCTAGACCTCGGC
GGCGCACTCAACACCCTGCTCGCCCGCCACGGACTACCTGCAAAAAGCATGGACGAAATC
CGCACCCAAGCCAGCCACGGCGCGGCAGGACTGATCAAGCTCGGCGCAGGCATCACCCCC
GACCATCCCGACTATGCCCGATGGCGCACCGAATACCTTGACGAATACGACAGCCGCTAC
GCCCAAGACACCACCCTCTTCGACGGCGTAAACGAACTCATCGCCGAACTCGGAAAACGC
GGCATCAAATGGGGCATCATCACCAACAAACCCATGCGCTTCACCGACAAACTCGTCCCC
AAACTCGGCTTCATCATCCCACCCGCCGTCGTCGTCAGCGGCGACACCTGCGGCGAGCCC
AAGCCCAGCGTCAAACCCATGCTGTATGCGTGCGGACAAATCCACGCCGACCCGCAACAC
ACACTCTACGTCGGCGACGCGGAACGCGATATACAGGCGGGGCGCAACGCCGGTATGACG
ACCGTCCTCGCCGAATGGGGCTACATCGCTCCCGAAGACGATACCGGCTCATGGCAGGCG
GATTTCCACATCCGCACGCCACTCGATCTGCTCGAATGTCTGGACAAAATACAGCCCTGA
AAAATATCCGCCCCACAAACATATAGTGGATTAACAAAAACCAGTACGGCGTTGCCTCGC
CTTAGCTCAAAGAGAACGATTCTCTAAGGTGCTGAAGCACCAAGTGAATCGGCTCCGTAC
TATTTGTACTGTCTGCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATAAAAC
TGCCGTCTGAAACCTGATTTCAGACGGCAGTTCCGCCTTCAAACCGAATCAAAGCCCGTC
AAAACCTGCGTTTGAGCTTGCACGCCTGAAGGATGTGTACCGCCAATTCCTCAACCGACT
TATCCGTCGTATTCGCAAACGGAATCCCATGCCGTCTGAACATACTCTGCGCGTCCGCCA
CCTCGCTGCGGCATGTATCGATTTTGGCATAAGTTGAATTCGGGCGGCGCTCTTGGCGGA
TGGCCTGCAAACGTTCCGGCTGGATGGTCAACCCGAACAGCTTATCCCTATAAGGCTTGA

CCATACGCGGCAGATCGGCCGATTCCAAATCGTCGGGAATCAGCGGATAGTTTGCCGCAC
GGATGCCGTATTGCAGGGCGAGGTAAAGGCAGGTCGGCGTTTTGCCCGAACGCGATACAC
CCATCAAGATTACATCCGCTTCCTGAAGGTTCTTATCGCTGACCCCGTCGTCGTGGTTCA
AAGAAAAATTGACCGCTTCCATACGCGCATCATAACGCTTCGTATTACCGATACTGTGAT
GCCCCTGCCCGGATGCCGTGGCTTCGGTATTGAGTTCCTTCTCCAACAGTCCCAAAAAAG
TCTCAAAGAAATTAATCTGAAAAGCATCCGCCCCTTTGATAATCCGACGGATTTCGTCAT
CAACCACACTGACAAACGCAATCGGACGCTGACCGTTTTCCTGCCGGCTCCGATTGACCT
TCTCCACCACCGCGCGCGCCTTTTCCGGCGTATCGACAAACGGATGCGTATGGCGTTTGA
ACGACAGATTGCCAAACTGGTTCAGCAACGCCTCGCCGATATTCTCAGCAGTCAGACCGG
TACGGTCGGAAATGTAAAACACATGGCGCGGACTGCTCATCTTCCATCCTTAAAACACAG
GTTTAAAATCCCATCATAACAGCAGCAGCAGACACAAGGAAAGCACCCGCAGCACACCTA
CCTCGGATTACACCCAAACAGACACAATATAATTTTGAAATAAAATTATTTATATAAAGT
ATTTTTTGGCAGAAAATTTTAAAAAATAAACAAAAAAATCAAACAGAAAAAACATTAACC
TATTCAAACCACCTGTTTTACAAAGAAAATACCCAAAAAAAGAAGTATACCGGCTGTAAG
TTTCAAACCGCTACACACGCCGAAACCGCAATTTTTCAGACGGCATCATGATTTTAAAAC
GGATAAAACACATGACAGCAGAGGAACGAATCGCTTAAAATAAGCACGCGGATTTGTTTC
TTTTTTAACATATTTTGGATTGGACACACAAATGGCCGACAACTACGTAATCTGGTTTGA
AAACCTGCGTATGACAGATGTTGAACGCGTGGGCGGTAAAAACGCCTCGCTGGGCGAAAT
GATCAGTCAGCTGACCGAAAAAGGCGTTCGCGTCCCCGGCGGCTTTGCCACCACGGCCGA
AGCCTACCGCGCATTCCTCGCACACAACGGTCTGAGCGAACGCATTTCCGCCGCACTGGC
AAAATTGGATGTCGAAGACGTTGCCGAACTGGCACGCGTCGGCAAAGAAATCCGCCAATG
GATTTTGGATACGCCTTTCCCCGAACAGCTCGATGCCGAAATCGAAGCGGCATGGAACAA
AATGGTTGCCGATGCCGGCGGTGCGGACATTTCCGTTGCCGTACGTTCTTCCGCAACTGC
CGAAGACCTGCCGGACGCATCATTCGCTGGACAACAGGAAACCTTCTTGAACATCAACGG
CTTGGATAACGTTAAAGAAGCGATGCACCATGTATTCGCTTCCCTGTATAACGACCGTGC
CATTTCTTACCGTGTCCACAAAGGCTTCGAACACGACATCGTCGCCCTTTCCGCCGGCGT
TCAACGCATGGTGCGTTCCGACAGCGGCGCATCAGGTGTGATGTTCACCCTCGACACCGA
ATCCGGCTACGATCAAGTCGTCTTTGTTACCTCCTCTTACGGTCTGGGCGAAAACGTCGT
ACAAGGTGCGGTCAACCCGGACGAATTTTATGTGTTCAAACCCACGCTCAAAGCGGGCAA
GCCCGCCATCCTGCGTAAAACCATGGGTTCAAAACACATCAAAATGATTTTTACCGACAA
AGCAGAAGCCGGTAAATCCGTAACCAACGTCGATGTCCCCGAGGAAGACCGCAACCGCTT
CTCCATTACCGACGAAGAAATTACTGAGTTGGCGCATTACGCACTGACCATCGAAAAACA
CTACGGCCGCCCGATGGATATCGAATGGGGACGCGACGGCTTGGACGGCAAACTCTACAT
CCTGCAAGCCCGTCCCGAAACCGTAAAATCCCAAGAAGAGGGCAACCGCAACCTGCGCCG
CTTCGCCATCAACGGCGACAAAACCGTATTATGCGAAGGCCGCGCCATCGGTCAGAAAGT
CGGTCAGGGCAAGGTGCGCCTGATTAAAGATGCTTCCGAGATGGATTCCGTCGAAGCCGG
CGACGTACTCGTTACCGACATGACCGATCCGGATTGGGAACCCGTGATGAAACGTGCTTC
TGCCATCGTTACCAACCGCGGCGGCCGTACCTGCCACGCCGCCATCATCGCGCGTGAATT
GGGCATTCCTGCCGTTGTCGGCTGCGGCAATGCAACCGAATTGCTGAAAAACGGTCAAGA
AGTTACCGTATCCTGTGCCGAAGGCGATACCGGCTTTATCTATGCCGGTCTGTTGGACGT
ACAGATTACCGATGTCGCCTTAGACAATATGCCTAAAGCACCTGTAAAAGTCATGATGAA
CGTCGGCAATCCCGAACTCGCATTCAGCTTCGCCAACCTGCCCAGCGAAGGCATCGGCTT
GGCGCGTATGGAATTTATCATCAACCGCCAAATCGGTATCCACCCCAAAGCCTTGTTGGA
ATTTGACAAACAAGACGACGAATTAAAAGCGGAAATTACCCGCCGTATCGCCGGTTACGC
GTCCCCTGTCGACTTCTACGTCGATAAAATCGCCGAAGGCGTGGCGACATTGGCCGCATC
GGTTTATCCGCGTAAAACCATCGTCCGTATGTCCGACTTCAAATCCAACGAATACGCCAA
CCTGGTCGGCGGCAACGTATACGAACCGCATGAAGAAAACCCGATGTTGGGCTTCCGTGG
TGCGGCGCGTTATGTCGCCGACAACTTCAAAGACTGTTTCGCCTTGGAATGCAAAGCCTT
GAAACGCGTCCGCGATGAAATGGGGTTGACCAACGTTGAAATCATGATTCCGTTCGTCCG
CACTTTGGGCGAAGCCGAAGCCGTTGTCAAAGCCCTGAAAGAAAACGGCTTGGAACGCGG
CAAAAACGGCCTGCGCCTGATTATGATGTGCGAGCTGCCGAGCAACGCGGTATTGGCGGA
ACAATTCCTGCAATACTTCGACGGCTTCTCCATCGGCTCGAACGACATGACCCAACTGAC
CCTCGGTCTCGACCGCGACAGCGGCTTGGTATCCGAATCGTTTGACGAACGCAACCCTGC
CGTCAAAGTGATGCTGCACCTTGCCATCTCCGCCTGCCGCAAGCAGAACAAATATGTCGG
CATCTGCGGTCAAGGCCCGTCCGACCATCCGGACTTCGCCAAATGGCTGGTTGAGGAAGG
CATTGAAAGCGTTTCCCTGAACCCGGATACCGTCATCGAAACTTGGCTATATTTGGCGAA
TGAATTGAACAAATAATCAATGCCCATACCCCCGAGCCTGAAAAGCGCGGGGGTATTTTT

TTCCAAGCAGCTCCGCTCAGACGGCATTTCCTGCCGATGCCCCGTCCGCGATAATATTG
ACACCCACGCGCCGACTGCCTACAATTCCCCCTTCCCCGAGCAACCGGCAACGGTCAGCT
TCTTCTTTCAGACGGCATCTGCCTGTCTTTTTCCTCTTCAAAATACATCATTATTATGCA
CGTCTCCGAATTACAAACCCTGCACATTTCCAAACTCTTAGAATTGGCGGAAGAACACGG
CATCGAAAACGCCAACCGATTCCGCAAACAAGACCTCGTATTTGCCATCGTCCGCCAGAT
GATGAAAAAAGGCGAGGGTTTCACCTGCTCCGGCACGCTTGAAATCCTGCCCGACGGCTT
CGGCTTCCTCCGCAGCGCGGACACGTCCTATCTTGCCGGCCCCGACGACATCTATGTCTC
GCCCACCCAAATCCGCCGCTTCAACCTGCATACGGGCGACACCATCGAAGGAAGCGTGCG
CGTCCCAAAAGACAACGAACGCTATTTTGCCCTGGTCAGGCTTGATACCATCAACGGCGA
CCACCCGGAAGTATGCCGCCATAAAATCCTGTTTGAAAACCTGACCCCGCTGTTTCCGAC
CGAACAGTTGAAGCTGGAACGCGACTTAAAGTCCGAAGAAAACCTGACCGGACGTGCCAT
CGACCTGATTTCCCCTATCGGCAAAGGTCAGCGCGCCCTCTTGGTTGCCCCGCCCAAAAG
CGGTAAAACCGTGATGCTGCAAAACATTGCCCACGCCGTTACCGCAAACTATCCCGAAGT
CGAACTCATCGTCCTCTTGATTGACGAACGTCCCGAAGAAGTAACCGAAATGAGCCGCTC
CGTCCGTGGCGAAGTAGTCTCCTCCACCTTTGACGAGCCGGCTCAACGCCACGTCCAAGT
TGCCGAAATGGTGCTTGAAAAAGCCAAGCGTATGGTGGAACACAAAAAAGACGTGGTCAT
CCTGCTGGATTCGATTACCCGCCTTGCCCGCGCCTACAATACCGTCGTGCCTACCTCGGG
CAAAATCCTGACCGGCGGTGTCGATGCCAACGCGCTGCATCGTCCCAAACGTTTCTTCGG
CGCGGCGCGCAACGTGGAAGAAGGCGGTTCGCTGACCATCATCGCCACCGCATTGGTTGA
AACCGGCAGCCGTATGGACGATGTGATTTACGAAGAATTCAAAGGCACCGGCAATATGGA
ATTGCACCTTGACCGCCGTATGGCGGAAAAACGCCTCTTCCCCGCCATCAACATCAACAA
ATCCGGCACGCGCCGCGAAGAGCTGCTTGTCCCAAACGACCAGTTACAACGTATGTGGCT
CTTACGCAAGTTCCTGCACCCGATGGACGAAATCGAGGCAGCCGAATTTTTAATCGGGAA
AATCAAAGCCTCTAAAAACAACGACGATTTCTTTGAACTGATGCGCGGCAAATAAACGCG
CCGCCCGCATTAATGCGAAATGCCGTCTGAAGCCTGAAAATCGGGTTTCAGACGGCACTT
TCATTCACACGGTCGGCGCAGCTCCTCCCTCCCCCCGTTAACGGCGCAACCGTCGGCAGT
GTCCGTGTCCGCTTGCCGAAAGCGCGGCCTTTGCAAAGCCGGCTTGAACGCATCCGTACC
GGCAATGAAAACCGATGCGGGAACTTGATTCAAGGTTGCGCCGAACCGCACTCATTTTTG
TATAAATTTGGGGCTGTCCTAGATAACTAGGGAAATTCAAATTAAGTTAGAGTTGCCCCT
ATGAGAAAAATTCGTCTAAGCCGGTATAAACAAAATAAACTCATTGAACTGTTTGTCGCA
GGTGTAACTGCAAGAACAGCAGCAGAGTTAGTAGGCGTTAATAAAAGTACCTCAGCCTAT
TATTTTCATCGTTTACGATTACTTATTTATCAAAACAGTCCGCATTTGGAAATGTTTGAC
GGCGAAGTAGAAGCAGATGAAAGTTATTTTGCTGAACGACAAAACCATATCAATGGAATT
GGGAACTTTTGGAACCGGGCAAAACGTCATTTACGCAAGTTTGACGGCATTCCCAAAGCG
CATTTTGAGCTGTATTTAAAGGGGTACGAACGACGTTTTAACAACAGCGAGATAAAAGTT
CAAATTTCCATTTTAAAACAATTAGTAAAATCGAGTTTATCCTAGTTATCTAGGACAGCC
CCATAAATTTTTATAGTGGATTAACAAAAACCAGTACGGCGTTGCCTCGCCTTAGCTCAA
AGAGAACGATTCTCTAAGGTGCTGAAGCACCAAGTGAATCGGTTCCGTACTATTTGTACT
GTCTGCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATATTTCCAATTAAAAA
TTTTTAATATATTAATCAATAAATTAATTTTATAAAATAAAAATATTGTCAACAATATTT
TGCCTTATCGCCCAAACCTCTGTATATTTTCCTACAGTAAATTGTTGACAATCCATACGC
CCACATATGCGCCGCCTAAGGATAAATCCTCCCGCCGGACAACGGGTGCAAGGGATCGGA
TGCGATATTTCCATATTCAAACAAGGGATTGGCGGCACATCGGCAAAATCCCCGCGCCGC
CCCGGTCCGGCAGGGCTTGCGTCCCTCCCGGACAAGCCCCGACCCCGCCTTTCCGAAAGA
CGGGCTCAACCATTAAGGAAACTTTAGTCAAAATGAAAAAACACATATGGGCGGCATCTT
TGCTGCCGGCATCCCTATCGGCAGAACCTTTAAACTGGTGGAAGCCTTATTCCGCCGTCA
ATTCGGGCGATACCGCCTGGGTGATGACTGCGGCTGCCTTGGTACTGTTGATGACGCTTC
CCGGGCTGGCTTTATTCTACGGCGGTATGGTGCGGAAAAAAAACCTGCTCTCGACGATGA
TGCACAGCTTTTCCATCGCGACATTGGTGGGCATCCTTTGGGTCGCCGTCGGCTATTCTT
TAGCGTTCACGCCGGGAAATGCCTTTATCGGCGGTTTGGGGCGCGTATTTTTAAGCGGGA
TGCAGATAGACGCTACCGCACAGATGCTGACCGTGTCGCCCAATGCGCCGACTGTTCCCG
AACCGGTATTTATGTTTTTTCAGATGACGTTTGCCATTATTTCGACCGCCATTATTACCG
GCGCGTTTGCCGAACGGATGAAATATTCGGCAATGATGCTGTTTTCGGGCATATGGTTTT
TATTGGTTTATGTGCCGGGCGCGCATTGGGTGTGGGGCGGCGGCTTTATGAGCAAGGGCG
GCGTATTGGATTATGCCGGCGGTACGGTGGTGCACATCAATGCCGGTATCGCGGGACTCG
TCGCCGCCTTGGTTTTGGGCAGGCGCATAGGCTACGGGCGCGAGGCGATGCCTCCGCACA
ATATGGCGATGACACTGATCGGCGCGGCAATGTTGTGGTTCGGCTGGTTCGGCTTTAACG

CCGGATCGGCGCTTGCGGCAGACGCGGCGGCGGTATGGCGATGGCGGTAACGCAGGTGT
CGGCCGTATTCGGCGCGGCAGGCTGGCTTGCCTGCGAAAAAATAGCGGGACACAAACCTT
CCGCTTTGGGGCTGGCTTCCGGCGCGGTTTCCGGTCTGGTCGGCATCACCCCTGCCGCCG
GCTTTACCGGCCCGTCGGGCGCGGCCGCCATCGGTATATTGACTGCCGCCGCGTGCTTTG
TGTCCGTCACCGTCGTCAAACACAAATTGCGTTACGATGATTCTTTGGACGCTTTCGGCA
TACACGGATTCGGCGGGCTGGTGGGCGGAATATTGACCGGCATCTTTTTCGACAACCGCA
TTTTCGGCGGGGATGCGGCAGTTTGGCAGCAGTTGTGGATACAGGTAAAAGACGGGGTCG
TTATGGCGGCATACAGCGGGCTAATGAGTTGGGCGATTTTGAAGGTCGTGGGGAAAATCT
GCGGCGGCCTGCGCGTCGGCAAGGATGTCGAACGCGAAGGTTTGGATCTGAATATCCACG
GCGAACGCGTGGAATAAGGGCGGCTATGCCGTCTGAAGCCTGAAAATCGGGTTTCAGACG
GCATTTTTCACGTTTGCCGCCGATGGATAAACATATAGTGGATTAACAAAAATCAGGACA
AGGCGACGAAGCCGCAGACAGTACAGATAGTACGGAACCGATTCACTTGGTGCTTCAGCA
CCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCCGTACTGGTTTTTGTTAAT
CCACTATACTGCCTGCGACGCTTAACGGCTGTCTTTCCACTGATAATATTTCGAGCCGAG
GAAAGAACCGAGTTTGCGCAAAAACGGTTGCAGGATAATATTCGGCTGGCGCAACCGCTC
AAACGGCTTCAGACGGCATTCGTCCCCTAAAATTGCTTCGGCAACCGCCAGACCTGCAAT
GCCTGTTATCGCCATCCCGTGTCCGGAATAACCTTGCGCATAAAAAACATTCGGGGCTAA
ACGTCCGAAATGCGGGACAAGGTTGGCGGTAATGTCGCACTCCCCGCCCCACGAATATTC
GATTTTGACATCGGCAAGCTGCGGAAAAACTTTAAGCATATCTTGGCGGACAAGCTCGGT
CATACGCTCAGGATTGTCGATAAACTCGTTATCCTTACCGCCGAAAAGCAGTCTGCCGTC
CGCGCTGAGGCGGTAATAATCCAAAATATGGCGGTTGTCGCATACTGCCATATTGTTACG
GATAAGCCCTTTTGCGCGCGCCCCCAAGGGTTCGGTCGCAATAATAAAGGTGCTGACAGC
AATCGCCTTGCGTTCCAAAGGCCGGAATATCGGGTTCAAACCTGCATAAGTATTGACAGC
ATAGACCACATTTTTGCACTCGACGCTGCCTTCGGGCGTGTAAACCAGCCAACCGTTTTG
ATGCGGTTCGATGCACGTCATCGGGGATTGCTCGAAAATCTGCGCACCGGCTTCGGCAGC
GGCACGAGCGATGCCCAAAGTGTAAGTGAGCGGATGCAGGTGTCCGGATAAGGGGTCGAA
TTGTGCCCCTTGGTACATATCGCTGTCAAGCTGCTGTTTCAACTCGGCTTTATCCCAAAG
TTGATAATGACTCGCACCGTAATGCCGTTGGGCGTGTTCATGCCACTGCTGCAACTCTTC
CCAATGCTGCGGACGGACGGCAACCGTGGCATAACCGCGCTGCCAATCACAATCGACGGC
ATGTTTGCGGACGCGTTCGTCCACCAGTTCGACCGCCTGCAAAGACTGTTGCCAAAACCA
TTGCGCCTGCTCCAAGCCGACCTGTTTTTCAATTTCCCCCATACCGCAGGCGTAATCGCT
GATAACCTGCCCGCCACTCCGTCCCGACGCGCCGAAACCGATACGCGCGGCTTCCAACAC
AACCGTTTCATGTCCCTGCTCCGCCAAGGGCAATGCAGTGCACAAACCACCCAATCCGCC
GCCGATGATACAGGTATCGGTTTTCAGACGGCATTGAAGTTTCGGATAAACAGTATGAGG
ATTAACCGAACTGAAATAATAAGAAGGCAGATATTCTTGAAAATCAGGGCGAATCATTGT
GTTTGCTTTATCAGGTGTATTTTCGGACGGAATGATACAGGCTGTCGGGCCATATCGTCC
AAACAGAAAATCGGTTGAAGAAAACAGGCTGACCCAGTCATGCGGTCAGCCTGCCTTATT
AATTAATTTGCTTTCTCGGCAGCCAATTTTTCCTGGCGGTAGGCTTCTGCCGCTTCTCGG
TCACGCTTGGTTGCCTGCCTCATCATCCAATAATTGACGATGATGACCAATGTTCCGATG
ATGCCGATTAGGATGGTCGCCAAGACATTCATCTGAGGATCGAGACCCAACTTGATTTTG
GAGAAAATCACCTGCGGCAATGTGGATGAACCGGGGCCGGAGAGGAATGAGGTAATCACC
AAATCATCCAAAGACAGGGTAATGCCGAGCAGAAAGCCTGAAGCGATGGCAGGGGCAATC
AAAGGCAAAGTGATGACAAAAAAGATTTTCAGCGGGCGCGCGCCCAAATCCATTGCGGCT
TCTTCGAGCGACTGGTCAAGCTCAACCAGACGCGAACGGATAACAACGGTAATGTACGCC
ATACACAGCGTCGTATGTCCGAGGAAGATGGTGAAAAAGCCACGATCGAAGTAGAGATGT
TGTAACCATTCGCTGCCCTGCAAAAATATCTGTACCTGAATAATCAGCAGCAGCATAGAC
AGACCGGTAATCACGTCGGGCATCACCATAGGTGCGGAAATCATGCCAGCGAACAAGGTA
CTGCCGCGAAAACGTTTAATCCGCGCCATCGCATAGCCTGCCAGCGTGCCCAAAACGACG
GCGGCAAGCGAAGACACAACGGCAATCCGCAGCGACAGCCAAGCGGCTTCCAAGATGGTG
TCGTTTTCCAGCAATGCGCCGTACCACTTGGTCGAAAAGCCGCCCCAAACGGTTACCAGC
TTGGATTCGTTAAACGAATAGATGACCAAAACAACCAGCGGGATATACAGAAACGCCAGC
GACAGTGCCAACATCAGTTTCAAGAACCAAGATAATTTGGATTTCTGCATTATTTGGCTC
CTTCTTCCAATTCGCGGTTTTCATAATGCTGAAACAGGGCAATCGGCACGACCAGCAGCG
CGACCATCACGACGGCGACGGCGGAAGCCAGCGGCCAGTTGTTTTGATCGAAGAACGCCT
GCCACAAGACTTTACCAATCATCAGGTTTTCCGAACCGCCGACCAGCTCGGGAATGACGA
ACTCGCCGACAGCAGGGACGAAAACCAGCATGGAGCCTGCAATAATGCCGGTTTTCGACA
AAGGCAGGGTAATCGTCAAGAACGATTTGACCGGCCCCGCGCCCAAATCGGAAGCCGCTT

CAAGCAGGCGGTTGTCGAGTTTCACCAGTTGCGTGTATAGCGGCAGAATCATAAACGGCA
GATAGGCGTAAACCATCACCAAATTGAGCGAAAAGGCATTGTAGAACAAATCCAAAGGCT
CGCTGATAATACCCATTTTAATCAACAGGTTGTTTACAATGCCGTTATGCCCGAGCAGAC
CCATCCACGCATAGACGCGCAACAGGAACGATGTCCAAAAGGGCAGCATAATGGCAAGCA
GCAAACCATTGCGGACAGAAGGATTGGCACGAGAAATCGCATAGGCGGTCGGATAACCGA
CCAACAGACAAATTACCGTCGTAGTCAGCGCAGTCTTAATTGAAGACCAATAAGTCATCA
GATAGATATTGCTGTTTTCACCGTCGCCGAACGGATTGAGCGTACTCCAAAAAATTTTGGA
AGATGTCTGCATAATTTTGGTAGCTGACAGCAATATTCAGACGACCCAAATCCTCATCTA
TCGTCGTTAAAGGAGTAAACGGCGGGATGGCGATTTCTTGTTCGGCAAAGCTGATTTTCA
GCACGATGGCGAACGGAATCAGAAACAGCACCAAAAGCCAAATATACGGTACGGCAATCA
CCGCACGCTGCCCCGGACGGCGGAACAGTTTGTTTTTCAGTTTATTAAGGTTCATTGCAT
TCCCCTTAAATCAACGGAACAACGGAGTCGGTTGGTTTTCCGGCCAGCTGATATAGACGG
TTTCGTCCCAAGTCGGCGGTGTAATGTTGCGCACATACCAGTAAGGGGCGGGGACTTGGC
TTTTGACGACGCGCCCGTTGCCGAGCTTGATATGGTAAATGGCGAAGCTGCCCAAATAGG
CGATTTCTTTTACCGTGCCTTTCGCCCAGTTGTAGTCGCCCAAATATTCGGGTTTTTCTT
TATATAAATCAATATCCTCTGGTCGAATACTAACCCAAAGGTCCTGCTCGCTCGGACCAC
CCAAACCGTGATCGATGCGGACGTGGTTTTCCAAACCTTCGCATTCGATAACGGCATAGT
CGGCATGATCTTCAATCACCACACCGTCAAAGATGTTGGTTTCGCCGATAAACTCGGCAG
TGAAGCGGCTGTTGGGATAGTCGTACACGTCGCTGGGTGTGCCGACTTGCTGCAACTGAC
CGTCAGACATAATGGCGATGCGGGTCGCCATCGTCATCGCCTCTTCTTGGTCGTGCGTAA
CCATAATACAGGTTACGCCGACTTGTTCCAGCGTATTGACCAACTCAAGCTGGGTTTGTT
GGCGCAGTTTTTTGTCCAATGCACCGAGGGGCTCATCCAGCAGTAGAATTTTCGGACGTT
TTGCCAGACTGCGTGCCAAAGCAATGCGCTGCTGCTGACCGCCGGACAATTGGTGCGGTT
TGCGTTTAGCAAATTTGGTCATCTGAACCAGGCGGAGCATTTCTTCGACGCGCGCGGCGA
TTTCGCCTTTAGGCATTTTGTCCTGTTTCAGACCGAAGGCAATGTTTTGTTCTACGGTCA
TATGCGGAAAAAGCGCGTAACTTTGGAACATCATATTGATGGGGCGATCATAGGGTGCAA
GTTTGGTAATATCCTGACCATCAAGGATAATTTTTCCCTGATTGGGACTTTCCATACCCG
CCAGCATACGCAGCAGTGTAGATTTTCCGCTGCCGGAACTGCCCAAAAGGGCGAAGATTT
CGTGTTGATAAATGTCCAAGTCGATGTTATCGACAGCGTAATTGTCACCAAACTTTTTCA
CCAAACCTTGGATTTTGAGATAAGGTTTGGCTGAAGACGCAGTGGTTGCGGTCATAATGG
CAATACTCCAATAAAAAGACGAGTACCGGCAAAACGGATTTTCGAATGGGTGATAAAAAG
CTGTTTGATTGCTGGCGGGAGTTAAACGTTTGATGCCGTCTGAAACTCTTGTAAAGCGCA
CGGGCAGCATGAAATGGAACAAGATTCCAAAGAACTTTATATTATATTAGTTTATGCGGT
TTTCGGGCAATATAGTGGATTAAATTTAAACCAGTACAGCGTTGCGTTGCCTTGCCGTAC
TATTTGTACTGTCTGCGGCTTCGTCGCCTTGTCCTGATTTAAATTTAATCCACTATAAAT
ATGGATTTGAGCTTGTCGGAAAGCAACAGAAAAGAAAACACCGCCCATTTTTCTGGGCGG
TGTCGGAAAGCGTAATTATTTACGCAGACCCAAGCGGGTAATCAACGCGCGATACGTATC
GGGCTGGGTACGGCGCAAGTAGGCCAGCAGGCGGCGGCGTTGGCTGACCATTTTCAACAG
GCCGCGACGGCTGTGGTGGTCTTTGGGGTTGGCTTTGAAGTGGGGGGTCAGGTCGTTGAT
GCGGAAAGTCAACAGAGCGACTTGTACTTCGGAAGAGCCGGTGTCGCCTTCTTTGCGTTG
GAAATCTTTAACGATTTGTGCTTTTTGTTCTACGGTCAGTGCCATAATGAAAACTCCAAA
AATATAAGAATCCCCATAGGGATTCAGACAAGTTTGCCAAGCCTGAAGACAACGGCAAAC
TCCCTATGCCCAAGATAGGACAACGTGGCATTATGACACAATTCCCGCATTTCTGCACAA
TATTTTAAGACCGTGGAAGTAACCGTTAAAAATGCCGTCTGAAGCATTGTCTGCTTGAGA
CGGCAATGTTCAGACGGCATATGCGCTTCAAACCACGATTTCTTCTTTCTTCTTCGGTTC
TTTACCGATATTGTCGCGGCTCAAACCGAACATTAGCAGAAGCGGGCTGGCAACCAATAC
GGAAGAATAAATGCCGAACACGATGCCAATGGTCAACGCCATAGAAAAGCCGTGCAAGGC
CGCACCGCCGAACACCAGCATGGATACGACCATCGCCTCGGTCGAACCGTGGGTAATGAT
GGTGCGGCTCATCGTTGCGGTAATCGCGTTGTCGATGACTTCCGGCACGGCATGTCCGCG
CATCGCCGGCTTGCGGAAGTTTTCACGGATACGGTCGAAGACGACGACGGATTCGTTCAC
AGAATAGCCCAATACGGCAAGGATACCCGCCAAGACGGTCAGCGAAAATTCCCATTGGAA
GAAGGCAAAGCAGCCGAGAATAATCACGATGTCGTGCATATTGGCGATAATGGCAGATAC
GGCAAAACGCCATTCAAAACGCATCGACAGGTAAATAATGATGCCGATAACGACAAAACC
TAAAGCCATCAATCCATTACTTACCAATTCCTCACCGACTTGCGGGCCGATAAATTCGAC
TTGGCGCAAGGTAACGTCGGGACTGTCTTTTTTCAGCAAATCCATAACCTGATTGGACAA
CTGTGCGGAAGTAACACCTTCTTTGTTCGGCAGGCGGATCATGATGTGTTTGTTCGTACC
CAATGCCTGAACCTGTACATCACCTATTTTCAGCGTATCGAGGCGTTCGCGCATCTTATT

GACATCCGCACCCTGCTGATATTGGACTTCCATTACCGTACCGCCGGTAAATTCGACAGA
GAAATTCAGACCTCTGGTAACCAAAAAGAACACGGCAGCGATAAACGTAACCAACGAAAT
GAAGGTCGTCAGTTTGCCGTAGCTCATAAACGGAATATCGCGTTTGATTTTAAAGAGTTC
CATAGCTTACTCCTTGCCTCCTGCCATTTCGGCTTTCGGCTTCCACACCGAACCAATGGA
AATATTCTGCAATTTGCGTCTGCGTCCGTACCACAGATTGACCAACGCACGGAATACGAC
GACGGATGAATACATCGAAGTCAGAATACCCAAACAGTGTACGACCGCAAAACCGCGTAC
CGGGCCGGAACCGAATACCAAAAGCGCGATACCGGCAATCAGCGAAGTCAGGTTCGAATC
GACAATGGTCGCCCATGCGTGTTGGAAACCGAGATTGATTGCCTGCTGCGGCGGCACGCC
GGCACGCAATTCTTCGCGGATACGTTCGTTAATCAAGACGTTGGAGTCGATTGCCATACC
CAAAGTCAACGCCAGCGCGGCCATACCCGGTAACGTCAACGTTGCCTGCATGGCAGACAA
AATACCGATTAGGAACAGTATGTTGGCACTCAATGCAATGGTAGAAAAGAAACCCATCAG
ACGATAGTAAACCACCATGAATGCAGCAACGATGGCAAAACCCCATAAAGTCGAATGGAA
GCCTTTTTCGATGTTCTCCTTACCCAAAGACGGACCGATGGTACGTTCTTCGACAATCTG
CATCGGTGCGGCAAGAGAACCGGCACGCAACAGCAAAGACGTATCATTGGCTTCGGCTGT
CGTCATGCTTCCGGAAATTTCCACGCGTCCGCCGGTAATGGCAGTACGGATAACCGGCGC
GGTTACAACCTCGGATTTTCCTTGGTCGATCAAAACCATCGCCATGCGTTTGCCGACATT
TGCGGCAGTCAGTTCGCCGAAAATGCTGCCGCCCGCGCTGTCCAAGCTCAGACTGACGGC
AGGTGCGCCCATTTGGTCGAAACTCGGTTGCGCATCGTTGATGTTGTCGCCCGTCAGCTC
GACCTGTTTGCTGATCAGCAGAATTTCGGGACGATCTCCGCCGCTTGAAAGCAGCTCATA
ACCGCTCGGCACGTTGCCTTCCAATGCCTCGCGCAACTTGGCAGGATCGTCCTCCACCAT
ACGCAATTCCAAAGTCGCGGTACGGCCGATGATGTCTTTTGCCTTGGCAGTATCCTGAAC
GCCCGGAAGCTGCACGACGATACGGTCTGCACCGGACTGCTGGATGACGGGCTCGGCCAC
GCCCAACTCGTTCACACGGTTGTGCAGGGTAGTGATGTTCTGTTTGACCGCATCGGAACA
CACTTTATTGACCGCCTCTTCCGAAAGCGTCAAGACGATATTGCTGCCGTCTGAATTCAG
CGTTGCTTCAGaAACAGCTTGCGCAACTGCGGCAGAGCCTTTTGCACATCACCTGCATCC
TGCAAAGGGACGGTCAGGCTGTTTCCAGCCTGACGCACCGTGCCGCTGcGGATTTTTTCG
CGGCGCAGTTCGCGGCGGATGTCGCCCGAATAACGTTCAAACGTTTTCTGCATCGTTGCT
TTCATATCGACCTGCATGGTGAAATGCACGCCGCCGCGCAGGTCCAAACCCAAAAACATC
GGATTGGCTTTGATTTTCGCCATCCATTCGGGGCTGTCCGCCAACAGGTTGAGCGCGGTA
ATATACCCTTCGCCCAAAGTGTTTTCGATGACGTCGCGCGCTTTAAGCTGCGTTTCTGTG
TCTTTGAAACGCACTTTCAGTGAATTGTCCACAACAAACATCCCGTCGGTCTGAATACCT
GCGTTTTTCAGCGCGGCATCCACTTTGAATTGAGTCTGTTCGTTGATGATGATGGCTTGT
CGGTTGGTCGATACCTGCACGGCGGGTGTTTCGCCGAATAGGTTGGGCAGCGAATACACT
GCGGCAACCGCAATCGTGAACACAATCAGCAGATATTTCCATAAAGGATAACGGTTCATC
ATTGTTCCTTAATGGTTGGAACCCCACCCTTTCGGTGGTGTCGGAATCGGGCTATTTCAG
AAGAGGCAAAAACCCTTCCCAGCCAGGCAAGACCGGAAAGCGGCATCCTGAATATGCCGC
CCTGCGTGTCGGAACATGGTCAAGCCTTCGGTTGGAATTCAAAACAAAGTGCCGCATTCG
GGCTTTCCAGATGCGGCTTGTCGGCACAAATCAATCGACTTTTGCGGCAATCGCATTGCG
TTCCACTTCGACCTCGATTTTTGTACCCTGTCCGATATCCACGGTAAAAAACTGTTCGCC
GACTCTGGTTACCTTACCCTTGAAACCTGCCGCCAAGACCACTTTGTCGCCGACTTTCAA
GGCGGCAAGCATTGCCTGATGCGCTTTGAATTTCTTTTGCTGCGGACGCATGATCAGGAA
GTAGAACACCACCATAATCAACACTAAAGGAGCAAATTGTGCAACAGCTTGATTCATAAT
TTATCCGTTCTTTCTAATATGGTTGAAAATCGAGAGGGGTATATAATAACATAAGACCGT
AAACAATATATCGGGTTTGCCGTCCGTACCGACCGTATACCGCAGCCTGCCCGTCCACAA
ACCCATGTCCTTGACCCGCCTAATCTTGAAATTCTATGCACTGTTGCGCCTTTTTTTGGG
CAAAAACGCCCGCACCGCATGGATTTCGCATCCCGCCTGTGCCGGGCACGAACCCGGCGC
AAACCATCCCGATTCGCCCGACCGCATCCTCTGCATCGAGCAGGCATTGCGCCGCGCCGG
TATTTGGCAGCACCTCCAAACCATAGAGGCGGAAGAAATCAGCGATACGCGCCTCGCACT
TGTCCACTCGAGCAAATATCTGAACCGTTTGGAATCTTGCCTGCCCCAAAAAGGCAAGAT
TTCCCGCCTGGATAACGACACTGCAATCAGCACAGGATCGCTGTCTGCCGCACGCTTTGC
CGCCGGTTCGGCAGTTCAGGCAGTCGACATGGTCATGAACCGTAAAGCATGGCATGCCTT
TTGCGCCGCCCGCCCGCCCGGACACCATGCGGGCAGCGGCAAAGCCGGCGGATTCTGCCT
GCTGAACAACGTTGCCGCCGGCGTCATGCATGCCATTGCCGAATACCGCCTGAAACGCAT
TGCCGTCATCGATTTCGATGTCCACTACGGCGACGGTACGGCAGAAATATTCAAAGACGA
TCCGCGCATCCTGTTTTTCAACCTGTTTGAAACCGACCTTTTCCCCTTCCCCGAAAACAA
CGATATGCCCGACGGCGGCAATATGGTGCACCTGCCCTTGCCGCCAGGAACGGGCAGCCG
CACATTCCGCGAAGCCGTCCGCAGGCAGTGGCTACCCCGACTTGCCGCATTCAAACCCGA

ACTGGTGCTGCTGTCGGCAGGATTCGACGCACACCGTCTAGACGAATCGGGCAGGCTCAA
CCTGCACGAGGCGGATTTTGCCTGGCTGACACACAAAATCATTCAGACGGCATCGGGCTG
CCCCGGCAAAATCATATCCGTACTCGAAGGCGGCTACACCCTTGAACCTCTGGCACAATC
TGCTGCCGAGCATATCCGCGTGCTGGCAGGGCTGGGCAAATCCGATGCCGCAACCGCCTA
TCAGAAAACACTGGACCCAACAAAAAAACGGTTTGCCAAACCGAAAACCGGACAGGTGCG
CCAACCTACCCAATCCGACCGTTACGACATATAGTGGATTAACAAAAATCAGGACAAGGC
GACGAAGCCGCAGACAGTACAGATAGTACGGAACCGATTCACTTGGTGCTTGAGCACCTT
AGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACACCGTACTGGTTTTTGTTAATCCAC
TATAAATGCCGTCTGAAGCCGATTCGGTTTTCAGACGACATCATTATTTGACCTGACGCA
AACAGAAAGCCTTCCGCGTGCTTAAACACATCTCTTGAGACCTTTGCAAAAATAGTCTGT
TAACGAAATTTGACGCATAAAAATGCGCCAAAAAATTTTCAATTGCCTAAAACCTTCCTA
ATATTGAGCAAAAAGTAGGAAAAATCAGAAAAGTTTTGCATTTTGAAAATGAGATTGAGC
ATAAAATTTTAGTAACCTATGTTATTGCAAAGGTCTCCTCTTGTAGTAAAATATCTAAAC
AGATTCAACTGGAAACGGTCAAGCTTCTTCACGCCATTCCCATTTCAACCGCCGCCTGCC
AGATTCAGACGGTATTCCTACTTGAAAACAAGGAGCGTTACAATGAAAGCAGCACGTTTT
TACGACAAAGGCGACATCCGCATCGAAGACATCCCCGAACCGACCGTCGCCCCCGGCACT
GTCGGCATCAATGTCGCCTGGTGCGGCATCTGCGGTACTGACCTGCACGAATTCATGGAA
GGCCCGATTTTCATTCCGCCTTGCGGTCATCCGCACCCGATTTCCGGCGAGTCCGCACCC
GTAACGATGGGACACGAGTTCTCCGGCGTGGTCTATGCCGTCGGCGAAGGCGTGGACGAC
ATCAAAGTCGGCCAACACGTCGTGGTCGAACCCTACATCATCCGCGATGACGTACCGACC
GGAGAAGGCAGCAACTACCACCTCTCCAAAGATATGAACTTTATCGGCTTGGGCGGCTGC
GGCGGCGGTCTGTCCGAAAAAATCGCCGTCAAACGCCGTTGGGTGCATCCGATTTCCGAC
AAAATCCCGTTGGATCAAGCCGCTTTGATCGAACCGCTGTCTGTCGGACACCACGCCTAT
GTACGCAGCGGCGCGAAAGAAGGCGACGTCGCATTGGTCGGCGGTGCAGGTCCGATCGGT
TTGCTGTTGGCTGCCGTGTTGAAAGCCAAAGGCATCAAAGTCATCATCACCGAGTTGAGT
AAAGCACGCAAAGACAAAGCGCGCGAATCCGGCGTTGCCGACTACATCCTCGACCCGTCC
GAAGTCGATGTTGTTGCAGAAGTGAAAAAACTGACCAACGGCGAAGGCGTGGACGTGGCA
TTTGAGTGCACCAGCGTCAACAAAGTGTTGGATACTTTGGTCGAAGCCTGCAAACCTGCC
GCCAATTTGGTTATCGTATCCATCTGGAGCCACCCCGCCACCATCAACGTCCACAGCGTC
GTGATGAAAGAGTTGGACGTGCGCGGCACGATTGCCTACTGCAACGACCACGCCGAAACC
ATCAAACTGGTCGAAGAAGGCAAAATCAACCTTGAGCCTTTCATCACCCAGCGCATCAAG
CTGGACGAGCTGGTTTCCAAAGGCTTCGAGCGTCTGATTCACAACAACGAATCCGCCGTT
AAAATTATTGTGAGTCCAAACCTGTAAGCAAATTACATTTGCAACAGCACAAATGCCGTC
TGAAACGCTCAGACGGCATTTTCCAACGAAGCGCAACAGGCGTATTGCCCGAGAGCAGCC
GTTTAGCCTATCATTCCGAACAAACAGAAACAAACAGAAGGAAAATCATGGGAGATTCCG
TACTATCCGCCATCCAACAAACCATCACCCAGCGAAAATCTGCCAATCCGTCCGAATCTT
ACGTCGCACAGCTCTTGCATAAGGGCGAAGACAAAATCCTAAAAAAAGTGATTGAAGAGG
CGGGCGAAGTGTTGATGGCATCCAAAGACAAAAACCCGTCCCACCTGGTTTACGAAGTTG
CCGACTTATGGTTTCACACCATGATTCTTCTGACACACCACGACCTGAAGGCGGAAGACG
TATTGGACGAACTTGCGCGCCGTCAGGGGCTGTCGGGGCTGGTCGAAAAAGCCGCTCGCA
CAGAATCTTGAATTTATATTAAAATCCGCACTTTCCCACATTCAATCCGTCTGACCGCTT
TTCAGACGGCATCGGAGCCGTTATGGACAACTGTATTTTCTGCAAAATCGCCGCCAAAGA
CATTCCGGCGCAAACCGTCTATGAAGACGGCGAAATGGTTTGTTTCAAAGACATCAACCC
CGCTGCTCCGGTTCATCTGCTGCTGATTCCCAAAGTCCATTTCGATTCGTTGGCACACGC
CGCGCCCGAACATCAGATGCTGCTGGGCAAAATGATGTTGAAAGTTCCCGAAATCGCCAA
AGCGGCAGGACTGGCAGACGGCTTCAAAACCCTGATCAATACCGGAAAAGGCGGCGGACA
AGAGGTCTTCCACCTGCATATACACATCATGGGTACACCCGTATAAACCGTTATTTCACA
ATCAACCCCTAATACTTACTTAAGGATACATCATGGGCAGTTTTTCTCTGACGCACTGGA
TTATCGTACTGATTATCGTCGTTTTGATATTCGGCACCAAAAAACTGCGCAACGTCGGCA
AAGACCTCGGCGGTGCGGTTCATGACTTCAAACAGGGGCTGAACGAAGGTACAGACGGCA
AAGAAGCCCAAAAAGACGATGTAATCGAACACAAAAAAGACGAAGACAAAGCGTAATTTA
TGTTTGATTTCGGTTTGGGCGAGCTGGTTTTTGTCGGCATTATCGCCCTGATTGTCCTCG
GCCCCGAACGCCTGCCCGAGGCCGCCCGCACCGCCGGACGGCTCATCGGCAGGCTGCAAC
GCTTTGTCGGCAGCGTCAAACAGGAATTTGACACTCAAATCGAACTGGAAGAACTGAGGA
AGGCAAAGCAGGAATTTGAAGCTGCCGCCGCTCAGGTTCGAGACAGCCTCAAAGAAACCG
GTACGGATATGGAAGGCAATCTGCACGACATTTCCGACGGTCTGAAGCCTTGGGAAAAAC
TGCCCGAACAGCGGACACCTGCCGATTTCGGTGTCGATGAAAACGGCAATCCGCTTCCCG

ATGCGGCAAACACCCTATCAGACGGCATTTCCGACGTTATGCCGTCCGAACGTTCCTACG
CTTCCGCCGAAACCCTTGGGGACAGCGGGCAAACCGGCAGTACAGCCGAACCCGCGGAAA
CCGACCAAGACCGCGCATGGCGGGAATACCTGACTGCTTCTGCCGCCGCACCCGTCGTAC
AGACCGTCGAAGTCAGCTATATCGATACTGCTGTTGAAACGCCTGTTCCGCACACCACTT
CCCTGCGCAAACAGGCAATAAGCCGCAAACGCGATTTTCGTCCGAAACACCGCGCCAAAC
CTAAATTGCGCGTCCGTAAATCATAAAGAGAGCAATCCGGTGTCCGAAACACAAAACGAA
CAACCCGTCCAACCGCTTGTCGAGCATCTCATCGAGCTGCGCCGCCGCCTGATGTGGACG
GTTGTCGGCATCTTAGTCTGCTTTTTCGGCCTAATGCCGTTTGCCCAACAACTCTATACT
TTTATCGCCGACCCGCTGATGGCAAACCTGCCCAAAGACACCAGCATGATTGCCACCGAT
GTCATCGCACCATTTTTCGTGCCGGTCAAAGTTACCCTGATGGCGGCATTTTTAATTTCG
CTGCCGCATACGCTCTACCAAATCTGGGCATTTGTCGCGCCCGCACTCTACCAAAACGAA
AAACGCCTGATTACGCCGCTCGTCCTCTCCAGCGTCAGCCTGTTTTTCATCGGCATGGCA
TTTGCCTACTTTTTGGTTTTCCCCGTCATTTTCAAATTCCTTGCCAGCGTTACCCCTGTC
GGTGTCAATATGGCGACAGACATCGACAAATACCTCTCCTTCATCTTGGGGATGTTTGTT
GCGTTCGGCACAACGTTTGAAGTCCCCATTGTCGTTATCCTGTTAACCAAAATTGGTGTG
GTAACAACCGGACAGCTCAAACGCGCCCGCCCCTATGTGATTGTCGGCGCGTTTGTCATT
GCCGCCATCATCACGCCGCCCGATGTGATTTCACAAACCCTGCTTGCCATTCCGCTGATT
CTCTTATACGAAGCAGGTATTTGGTTCGGACGCTTTTTCACGCCACGTTCAGAACAGGAT
GGCGACATACGGCCGCCTGCAACAACCTGACACTATGCCGTCCGAACCTCCGCCTCATAC
CGCAACAGATTAAGGAATACCTTTGAATACCCTCTATTTAGGTTCAAACAGCCCGCGCCG
AATGGAAATCCTGACACAGTTGGGCTACCGCGTCATCCAACTGCCTGCCGGCATCGACGA
ATCCGTTAAAGCCGGCGAAACACCTTTCGCTTACGTTCAAAGGATGGCAGAAGAAAAAAA
CCGAACCGCCCTGACCCTCTTTTGCGAAACCAACGGCACAATGCCCGATTTCCCCCTGAT
TACCGCCGACACCTGCGTCGTTTCAGACGGCATCATATTGGGCAAACCCCGCTCCCAAGC
CGAAGCAATCGAATTTTTAAACCGATTGTCCGGCAAACAACATACCGTCCTGACTGCTGT
CTGCATTCATTATCGCGGCAAAACGTCAAGCCGCGTCCAAACCAACCGCGTCGTTTTCAA
GCCCCTGAGTTCGGAAGAAATTTCCGCCTATGTGCAAAGCGGCGAGCCGATGGACAAAGC
CGGTGCCTACGCCGTACAAGGCATAGGCGGCATCTTTATCCAATCTATCGAAGGCAGCTT
CAGCGGCATTATGGGACTGCCCGTTTATGAAACCGTTTCGATGTTGCAGGATTTGGGATA
CCGCTCCCCCTTGTCCGCCCTTAAACCGTAAAGACCGCCGTGAACAGACAAACCGCTTAC
CTCCTTGCCTCTTTCAGCCTGATCGCACTGATAATCCTGTCCCTTTCCTGGGAACTGTGG
ATTGCACCGTTGCGCCCGGGCGGCTCGTGGCTCGCGCTCAAAGCCCTCCCCCTCTGTCTG
CCGCTTTCAGGCATCTTGAAAAAGAAAATCTATACTTACCAATACAGCTCCATGCTGGTT
CTGATTTACTTTGCCGAAGCCGTCATGCGCCTGTTCGACGCCTATCCCGCAGAACAGATT
TGCGCCGCGCTTTCCGCAGTATTCAGCATCATCTTCTTCATATCCTGCCTGTCGTTCGTC
AAACAATACAAGGAAACAAACAATGCCCGCTGAAACGACCGTATCCGGCGCGCACCCCGC
CGCCAAACTGCCGATTTACATCCTGCCCTGCTTCCTTTGGATAGGCATCGTCCCCTTTAC
CTTCGCGCTCAAACTGAAACCGTCGCCCGACTTTTACCACGATGCCGCCGCCGCAGCCGG
CCTGATTGTCCTGTTGTTCCTCACGGCAGGAAAAAAACTGTTTGATGTCAAAATCCCCGC
CATCAGCTTCCTTCTGTTTGCAATGGCGGCGTTTTGGTATCTTCAGGCACGCCTGATGAA
CCTGATTTACCCCGGTATGAACGACATCGTCTCTTGGATTTTCATCTTGCTCGCCGTCAG
CGCGTGGGCCTGCCGGAGCTTGGTCGCACACTTCGGACAAGAACGCATCGTGACCCTGTT
TGCCTGGTCGCTGCTTATCGGCTCCCTGCTTCAATCCTGCATCGTCGTCATCCAGTTTGC
CGGCTGGGAAGACACCCCTCTGTTTCAAAACATCATCGTTTACAGCGGGCAAGGCGTAAT
CGGACACATCGGGCAGCGCAACAACCTCGGACACTACCTCATGTGGGGCATACTCGCCGC
CGCCTACCTCAACGGACAACGAAAAATCCCCGCCGCCCTCGGCGTAATCTGCCTGATTAT
GCAGACCGCCGTTTTAGGTTTGGTCAACTCGCGCACCATCTTGACCTACATAGCCGCCAT
CGCCCTCATCCTTCCCTTCTGGTATTTCCGTTCGGACAAATCCAACAGGCGGACGATGCT
CGGCATAGCCGCAGCCGTATTCCTTACCGCGCTGTTCCAATTTTCCATGAACACCATTCT
GGAAACCTTTACTGGCATCCGCTACGAAACTGCCGTCGAACGCGTCGCCAACGGCGGTTT
CACAGACTTGCCGCGCCAAATCGAATGGAATAAAGCCCTTGCCGCCTTCCAGTCCGCCCC
GATATTCGGGCACGGCTGGAACAGTTTTGCCCAACAAACCTTCCTCATCAATGCCGAACA
GCACAACATATACGACAACCTCCTCAGCAACTTGTTCACCCATTCCCACAACATCGTCCT
CCAACTCCTTGCAGAGATGGGAATCAGCGGCACGCTTCTGGTTGCCGCAACCCTGCTGAC
GGGCATTGCCGGGCTGCTTAAACGCCCCCTGACCCCCGCATCGCTTTTCCTAATCTGCAC
GCTTGCCGTCAGTATGTGCCACAGTATGCTCGAATATCCTTTGTGGTATGTCTATTTCCT
CATCCCTTTCGGACTGATGCTCTTCCTGTCCCCCGCAGAGGCTTCAGACGGCATCGCCTT

CAAAAAAGCCGCCAATCTCGGCATACTGACCGCCTCCGCCGCCATATTCGCAGGATTGCT
GCACTTGGACTGGACATACACCCGGCTGGTTAACGCCTTTTCCCCCGCCACTGACGACAG
TGCCAAAACCCTCAACCGGAAAATCAACGAGTTGCGCTATATTTCCGCAAACAGTCCGAT
GCTGTCCTTTTATGCCGACTTCTCCCTCGTAAACTTCGCCCTGCCGGAATACCCCGAAAC
CCAGACTTGGGCGGAAGAAGCAACCCTCAAATCACTAAAATACCGCCCCCACTCCGCCAC
CTACCGCATCGCCCTCTACCTGATGCGGCAAGGCAAAGTTGCAGAAGCAAAACAATGGAT
GCGGGCGACACAGTCCTATTACCCgTACCTGATGCCCCGATACGCCGACGAAATCCGCAA
ACTGCCCGTATGGGCGCCGCTGCTACCCGAACTGCTCAAAGACTGCAAAGCCTTCGCCGC
CGCGCCCGGTCATCCGGAAGCAAAACCCTGCAAATGACCCGCGCCGGCGGATGCGGATAC
CGCCCGAAATGTAAAGCTCCATGCAAGACATTGCAAAAAACAGCAAACCGGTAGGGAAAT
ACGCTATCAAAAAACATTGCAGCCGTGTTAAGATAAACCGTCAAACAATCTTTTCACaGC
CCCGCCCGAAACAGGGTCGGGGCATACCCTTACGAAAAGGAAACACCATGAGCCGCGTAT
TACTCGTAGATGACGATGCCCTGCTAACCGAACTGCTGACCGAATACCTGAGCGCCGAAG
GTCTGAACGTCCGCAGCGTTCCCGACGGGGAAGCAGGCGTACAGGAAATCCTGAGCGGGC
AGTACGATGTAGTCGTATTGGATTCCATGATGCCCAAAATGAACGGCTTGGATGTCTTGA
AAAACGTACGCGCCCGAAGCACCGTCCCCATCATCATGCTGACCGCCAAAGGCGACGACA
TCGACCGAATCATCGGCTTGGAAATGGGCGCGGACGACTATGTCCCCAAACCCTGCACAC
CACGCGAACTCTTGGCACGCATCAATGCCATCCTGCGCCGCGCACAACACAGCGGCGAAC
AGAACAACGCACCCAACAGCATCTCCGTCAGCGATGTCGTCCTGTACCCCGCCAAACGCC
AGGCATCCGTCAAAGACATGCCGCTCGAACTGACCAGCACCGAATTCAACCTGCTCGAAG
TCCTGATGCGCCATGCCGGACAGGTAGTCAGCAAAGAAACCCTGTCCGTCGAAGCACTCG
ACCGCAAGCTGGCAAAATTCGACCGCAGTATCGACGTACACATCTCCAGCATCCGCCACA
AGTTGGGCGATGCCTCTCTGATTCAAACCGTACGCGGCTTGGGCTACCTGTTTGTCAAAA
ACTGAAATAAACAGATAAATGAAACTGTTCCAACGCATTTTCGCCACATTTTGCGCGGTT
ATCGTCTGTGCAATCTTTGTGGCGAGTTTTTCTTTCTGGCTGGTGCAGAACACCCTTGCC
GAAAACCAGTTCAACCAACGCCGCACCATCGAAACCACTTTGATGGGCAGCATCATTTCC
GCATTCCGGGCACGCGGGGACGCGGGTGCGCGCGAAATCCTGACGGAATGGAAAGACAGC
CCCGTCTCATCGGGCGTGTACGTTATACAGGGCGACGAGAAAAAAGATATCCTGAACCGG
TATATCGACAGCTATACCATCGAACGCGCCCGGCTTTTCGCCGCCGGACACCCGCATTCC
AACCTCGTCCATATCGAATACGACCGCTTCGGCGAAGAATACCTGTTCTTCACCAAAGAC
TGGGACAAACTCCAAGCCCGCCGCCTGCCCAGCCCCCTGTTGATCCCCGGCCTGCCGCTC
GCCCCGATTTGGCACGAACTCATCATATTGTCCTTCATCATCATCGTCGGACTGCTGATG
GCATATATCCTCGCCGGCAACATTGCCAAACCCATCAGAATCTTAGGCAACGGCATGGAC
AGGGTGGCAAACGGAGAACTTGAAACCCGTATCTCCCAACAGGTCGACGACCGCGACGAC
GAATTGTCCCATCTTGCCATCCAATTCGACAAAATGGTGGAAAAACTCGAAAAACTCGTT
GCCAAAGAACGCCACCTGCTCCATCACGTCTCCCATGAAATGCGTTCTCCCCTTGCGCGC
ATGCAGGCAATTGTCGGACTGATTCAGGCGCAGCCCCAAAAACAGGAGCAATATCTCAAA
CGGCTGGAAGGCGAACTGACCCGCATGGATACGCTGGCCGGGGAACTGTTAACCCTGTCC
CGTCTCGAAACTTCCAATATGGCTTTGGAAAAAGAAAGCCTGAAACTCCTGCCCTTCCTG
GGCAACCTGGTAGAAGACAATCAAAGCATTGCCCAGAAAAACGGACAAACGGTTACCCTG
TCTGCCGACGGAAAAATCCCCGAAAACACAACCATCCTTGCCAACGAAAGCTACCTGTAC
CGCGCCTTCGACAACGTCATCCGCAACGCCGTCAACTACAGTCCCGAAGGCAGCACCATC
CTGATCAACATCGGACAAGACCACAAACACTGGATAATCGACGTTACCGACAACGGCCCC
GGCGTGGACGAAATGCAGCTCCCGCACATCTTCACCGCTTTCTACCGTGCAGACTCCAGT
GCCAACAAACCCGGAACAGGACTGGGGCTTGCATTGACCCAACATATTATTGAACAGCAC
TGCGGCAAAATCATCGCCGAAAACATCAAACCGAACGGTCTGCGGATGCGCTTTATCCTG
CCCAAGAAAAAAACCGGTTCCAAAACAGAAAAAAGTGCGAACTGACCATAATACCGTCTG
AACCGGCTTCAGACGGCATTGCACAAACAGTTACCCCAAAACAAACACACAGCGGGTGTC
TCTCCAAAAATCACGCACATACCGTCCGCAAAGGAATATATCATGTCGGCACAAACCGAT
CCGGGCTACTTCTTCATGCCCAACCACATCATCCTGATAGGCGCGAGCGAACAACCGTAC
AGCCTGGGCGAACGTGTACTCAGCAACCTGCTGAGTACGCCCTTTCAAGGAAAAATCACC
CCCGTAAACCCGCGCCACCACACCATAGCCGGACTGCCCGCCTACACCAGCCTCAACAAA
ATCCCCGGCAATGCAGACCTGATTATTGCCGTTACCCTACCCGACAGTTACGACACCCTC
TTCAAAACCTGCCGCAAAAAGCAGCTCCGACACATCATCCTCATACAGGACTGGGACAAC
CTGTCTGCCGCAGAACTGCACACCGCCGAAACTGCCATCCGCAAACACCACGGCAACGGA
CTCAACATCACCGCCTGCACCACCGCAGGCATCCAACTGCCCTCACTCGGACTCAACATC
AGTACCCAAGACGGATACGCCGCAGGCCATACCGCCATACTGACCGGCAATGCCGCCGTC

AGCCGCCAAATCGACAACATCCTGAACAAACTCCGTCAAGGCACATCCCGCCACATCAGC
CTGCATCCCGGCATCAGCCCCATCACATCCGCCGATTGGCTCAACCGCTTCGGACACAGC
CTGCACACCAAAACCGCCGTCCTACACCACAACCCTGAAGAGGATCAGCGCAAACTGTTC
AGCGCAATCCGCCAATTTACCCGCCATACGCCGCTGATTCTCCACATCACCTGCCGCACG
ACAGAAACCGACCGTGCCGTACTGCACTGCCTCGCCCGACACTGCAACTTCCTCGTCAGT
TTCAACGCCGACGACCTCGAAGCCGCACTGCGCGCCCAACTGTCCGACCTTCCACCCCTG
TCCCGACTCGACATCCTGTCCGACACGCCTGCCGAATGGCTGCACGCGCACGCGCCAAAA
AACCTCACCCTCCACTTTCCCAACCTTCCCCACCACATCCGCAACGGACACCTGACCGGC
ACACCCACACCCTCAATCTGCCACGACATCGCCTCACGTCAGCTTGCCCACCCCGACACC
CAAGCCGTCCTAACCATCCTCAGTCCCTCCGGACACGAGGATTACAAAAAAACAGCACGC
GCCCTTATCCGCCTGTCCGAACAGACCGCCAAACCCCTGCTCGTCAGCAGCCCCTTTTCA
GACGGCATAACCCATTTCGACACCCCCACTCAGGCAATCCGCACCCTTTCCTACCGCAAC
ACCGCCGCCGCCCTGAAACAGGCACAGCTCGACATTGCACCGCCGCAGCCATGCCGTCTG
AAAACACCGCAACCCCAAAACATCAAAAAAGCCCTTGCAGCGGCAAACCCCTCCCTGCTC
GCCGAAGCCCTGCACCTCCCCCCCTACCGGCACACCACCCATAACGCCGTACAATTCCAA
TTCGACAGCCACCCCCTCTATGGCGACATCCTGACCGCACGCTGCAACGGACAAACCACT
GCCGTACTCCCGCCGTTTACCACGCTCGACAGCCGCCACCTTGCCCGCTTTGCCGAACTC
GACGGCACACAAACCCTCGACCAGTTCCTGCACACACTGACCGTCATTCCCGAATACCGC
CAACACATTCTCGGCATCACCCTCAACCTCAACGGCGGACAATACAGCAGCGATTTCCTC
TTAAAGACACCCGAAACACACGACACGCCCAAACGCAAGAACACAGGCAAAGCCGCCCAA
ACCCTCGAACATGCCGCCGCAAAAATGCAGAGTGCCGCCGCATACCTGAAACACAAAAAC
CCGACAGCCGCCGAATTTCTCCGCCACACAAGCGAAGCCGCCGCAGAACTGCTCGGCAGC
AAAACCGAAACCGGAGCAGCCGTACCCAACGTACTTGCCCCCTATCCCGCAGCATACCCC
AAAACACTGTCCCTAAGAAACAACACGACCGTTACCATTACCCCCATTTTGCCCGAAGAC
GCAGAAGCCAAACAGCAGTTCGTCCGCAGCCTCGGTCCCGAAGCACGGTACACACGCTTC
ATGACCCACACCAACGAACTGCCCGCAGCCACGTTGGCACGCCTGTGCAACCCCGATTAC
CACTGTGAAGCCGCATGGACGGCAAAGGATGCCGACAGCAACATCGTCGCCGTCGTCCGC
CACAGCCGCCTGAATCGCAACGAATGCGAATTCGGCATCACACTGGCGGAACATATGCGC
GGCAGCGGACTGGCACAGAAAATGATGGAACTCATCATCCAAACCGCCGCACAGCAAGGC
TACCGGACTATGAGTGCCGACATTCTCAAAACCAATACCCCCATGATCAAACTTGCTGAA
AAATCAGGATTTACCCTCAAGGAATCGGACACCGAAAAAAACCTGTACCGCGCATATCTG
AACCTTGCGGCAGACAAAACAACAGAAAAAACAAATAAAAACTTGCGCACCGACCACAAA
ATAACCTAAAATCGGCAGTTTCCATATATCCGATTTTCTCAAAAGGACTCAAAATGGTAG
TTATCCGTTTGGCACGCGGCGGCTCGAAACACCGCCCCTTCTACAACGTCATCGTTACTG
ACTCACGCAGCCGCCGCGACGGCCGCTTCATCGAACGCGTAGGCTTCTACAACCCCGTAG
CCAACGAAAAACAAGAGCGCGTCCGCCTCAATGCAGACCGCCTGAACCACTGGATTGCAC
AAGGCGCGCAAGTCAGCGACTCCGTTGCAAAACTGATTAAAGAACAAAAAGCCGCCTAAT
CCGCATTTGCCGCCATGACAGACACTCAAAACCGGGTAGCCATGGGCTACATCAAAGGCG
TATTCGGCATAAAAGGCTGGTTGAAAATTGCCGCCAACACCGAATATTCCGACAGCCTTT
TGGACTACCCCGAGTGGCATTTGGTCAAGGACGGCAAAACCATCAGCGTTACCCTTGAAG
CCGGAAAAGTCGTCAACGGCGAACTCCAAGTCAAATTCGAAGGCATAAACGACCGCGACT
TGGCATTCTCATTGCGCGGTTACACCATCGAAATACCCCGTGAAGCATTCGCCCCGACAG
AAGAAGACGAATACTACTGGACAGACTTGGTCGGCATGACCGTTGTCAACAAAGACCATA
CCGTTTTAGGCAAGGTAAGCAACCTGATGGAAACCGGCGCAAACGACGTATTGATGATTG
ACGGAGAACACGGGCAGATTCTGATTCCGTTCGTTTCCCAATATATCGAAACCGTCGATA
CCGGCAGCAAGACCATTACTGCCGACTGGGGTTTGGACTACTGATGCTTATCCAGGCAGT
TACCATTTTCCCCGAAATGTTCGACAGCATTACCCGCTACGGCGTAACGGGACGCGCGAA
CAGACAGGGAATCTGGCAGTTTGAAGCAGTCAATCCCCGAAAGTTTGCCGACAACAGATT
GGGCTACATCGACGACCGCCCGTTCGGCGGCGGCCCGGGAATGATTATGATGGCTCCGCC
GCTTCATGCGGCAATAGAACACGCCAAAACACAATCCTCCCAAGCTGCAAAAGTCATCTA
CCTCAGCCCCCAAGGGAAACCGCTTGACACACCAAAAAGCGGTAGAACTGGCAGAACTTC
CGCATCTGATTCTGCTGTGCGGACGGTATGAGGGCATAGACGAAAGGCTTCTGCAAAGCA
GCGTCGATGAAGAAATCAGCATCGGAGACTTCGTTGTTTCCGGCGGAGAGCTTCCCGCCA
TGATGCTGATGGATGCGGTATTGAGGCTCGTACCCGGCGTATTGGGCGATATGCAGTCTG
CCGAACAGGATTCGTTCTCAAGCGGCATTTTGGACTGCCCCCACTACACCAAACCCTTAG
AATTTCAAGGTATGGCTGTTCCGGAAGTATTGCGTTCCGGCAATCATGGCTTGATAGCGG
AATGGCGGTTGGAACAATCGCTGCGCCGCACCTTGGAGCGCAGACCCGATCTTTTGGAAA

AGCGCGTTTTAATCCCAAAGGAATCCCGCCTCTTAGAAACCATCCGGCAAGAGCAACGGG
AAATCCAATCATAATTTAGGAAAAAAACAATGAACCTGATTCAACAGCTCGAGCAAGAAG
AAATTGCCCGCCTGAATAAAGAAATCCCCGAATTCGCACCGGGCGACACCGTAGTCGTAT
CCGTACGCGTCGTGGAAGGTACCCGCAGCCGTCTGCAAGCCTACGAAGGCGTGGTTATTG
CCCGTCGCAACCGTGGTCTGAACAGCAACTTCATCGTCCGCAAAATCTCCAGCGGCGAAG
GTGTTGAACGTACTTTCCAACTGTACTCTCCGACCGTCGAAAAAATCGAAGTCAAACGCC
GTGGCGACGTACGCCGTGCCAAACTGTACTACCTGCGCGGCCTGACCGGTAAAGCTGCAC
GCATCAAAGAAAAACTGCCTGCACGCAAAGGTTGATTCAAACCGTTTTCCCCCAATGCCG
TCTGAACCTTCAGACGGCATTTCTTATTGCTGCGATGCCGGCAAGTCAACGATATGCCTG
CCCCTCATCCACCTTGATTAACAATCACTCCCGCCTATTCAAGCATAATTTACAGACCAA
ACGTTATACAGTATCATATCAGCCATATAACGATAAACGCTTTACCGCCTTCCCTTTCAA
AACCGCAAATCATCATGACACCTTCCCTTTTACTATCAGGATTGACCTTCCGCCTCATCC
TTGCCCTGATTGCCGTATCCCTTTTATGGGGCGTTTACTTTTGGGCAGTATCCGCATGAG
CATCATTGTCGAAAACCTGACCGTCAGcTACCGCCGCCGACCTGCCGTGCACCATGTGGA
CATTACTTTTGAAGAACATAGTATGTGGGCGGTTTTCGGTCCCAACGGCGCAGGGAAATC
CACCTTTCTCAAATCCTTGATGGGATTGCAGCCTATCGATACAGGCAGCATCCGGCTGGA
CGGATTGACCCGTCAGAACATCGCCTACCTTCCCCAGCAGTCCGATATCGACCGCTCCCA
GCCTATGACCGTTTTCGACTTGGCGGCAATGGGGCTATGGTATGAAATCGGCTTTTTCAA
AGGGATAAATACCGCTCAAAAACAACGCGTTCACGAAGCCTTGGAGCGCGTCGGAATGCA
ACGGTTTGCCGACCGTCAGATTGCCTATCTCTCAAACGGACAATTTCAGCGTGTCCTTTT
TGCCCGAATGCTGGTTCAAAATGCCAAATTCCTGCTGCTCGACGAACCCTTCAATGCCGT
TGATGCACGGACAACCTACGAGCTTCTCGACGTATTGCAGAAATGCCATTGCGGCGGACA
CGCCATCATCGCCGTACTGCACGATTACGAACAAGTCCGTGCCTACTTTCCCAATACCCT
GCTGCTCGCCCGCGAAAAAATTGCGGCAGGCGCAACCGAGACCATTCTGACAGAACCCTA
CCTCGCCCAAGCCAACGCCAAAATGCAGCAACAGGAAAGCCCCGACTGGTGCGCCTCATA
AATGCCGTCTGAAACCGAAAAaCCATGAATCTCTACGACCTGCTCCTTGCCCCCTTTGCA
GAATTCGACTTTATGCGCTACGCCCTCGCATCCGTCTTCTGCCTGTCCCTCAGTGCCGCA
CCCGTCGGCGTATTCCTCGTCATGCGCCGTATGAGCCTGATAGGCGACGCATTGAGCCAC
GCCGTCCTGCCCGGTGCCGCCGTCGGCTACATGTTTGCCGGCTTGAGCCTGCCCGCCATG
GGTTTGGGCGGCGTAGCCGCAGGCATGCTGATGGCACTGCTTGCCGGACTCGTCAGCCGC
TTCACCACCCTGAAAGAAGATGCCAACTTTGCCGCCTTTTATCTCAGCAGCCTCGCCATC
GGCGTAGTCCTCGTCAGCAAAAACGGGAGCAGCGTCGATTTGCTCCACCTCCTTTTCGGC
TCTGTACTTGCCGTCGATATTCCTGCCCTGCAGCTCATCGCCGCCGTCTCCAGCCTCACG
CTCATTACCCTTGCCGTCATCTACCGCCCGCTCGTACTCGAAAGCATCGACCCCCTGTTT
CTCAAATCCGTCGGCGGCAAAGGCGGGCTTTGGCACGTCCTCTTTCTCGTCCTGGTCGTC
ATGAACCTCGTATCCGGCTTTCAAGCCCTCGGCACACTCATGTCCGTCGGACTCATGATG
CTGCCAGCCATTACCGCCCGCCTGTGGGCGAAGCATATGGGCGCACTCATCCTCCTATCC
GTTCTGACAGCCCTGCTGTGCGGCTTGAGCGGACTGCTCATTTCCTACCACATCGAAATT
CCTTCCGGTCCCGCCATCATCCTCTGTTGCAGCGTCCTTTATCTCTTTTCCGTCATACTC
GGCAAAGAAGGCGGCATTCTGACCGG

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 6>:

gnm_6

GTTATGATATGTTTCGATATGTAAACATTTATAGGTTTGGAGCGATAAACAATGAATGCA
GTCGTAGTTGCCGTAATCGTTATGCTGGTGCTGTCGCTGTCGCGCGTGCACGTGGTATTG
AGCCTGACGGTCGGCGCGTTTGTCGGCGGCGCGGTGGCGGGTATGCCGCTGCAAAACATT
GCCGATGCGGCGGGACAGGTCAGTCAGGCGGGGATTATCCCCGTGTTCAACAAAGGTTTG
GAAGGCGGTGCGAAGATTGCGCTTTCTTATGCGATGCTCGGCGCGTTTGCAATGGCGATT
ACCCATTCCGGCCTGCCGCAGCAGCTTGCCGGCGCGGTCGTCCGCAAGCTGAACCGGGGC
GGTATGCCCGACAGCGTGCGTTCGGGCGAGGGCGCGGTCAAATGGCTGCTGCTTTCCATC
ATCCTTGTGATGGGCATGATGAGTCAGAACATCATCCCCATCCACATTGCCTTTATCCCG
ATGATTGTTCCGCCGCTGCTTTTGGTGTTCAACCGCCTGAAAATCGACCGCCGCCTGATT
GCGTGCGTCATCACTTTCGGGCTGGTTACGACTTATATGTTCCTGCCTTACGGCTTCGGC

GCGATTTTTTTGAACGAAATCCTGTTGGGCAACATCCATTCCGCCGCGCCGCAGCTTGAT
GTGAAAAACATTAACGTGATGGCGGCAATGGCGATTCCCGCGTTGGGAATGCTGGCCGGA
CTCCTGCTGGCGTTTGTCCATTACCGCAAACCGCGCCTGTACCAAAGCAACAATGCCGAT
ACGGCGGGCAACGCCGATGCGGCAAACCGTCCGCAGCCGTCCGCCTACCGCAGCCTGGCC
GCCGCCGTCGCCATTGCCGTATGCTTTGCCATCCAGTTGATGTATGAAGACTCGCTGGTG
TTGGGCGCGATGCTCGGTTTCGCCGTATTTATGATGTTGGGGGTCATTAACCGCGACAAG
GCAAACGACGTATTCGGCGAAGGTATCAAGATGATGGCGATGGTCGGCTTCATTATGATT
GCCGCGCAGGGTTTGCCGCCGTGATGAATGCGACCGGGCATATTCAGCCGCTGGTGGAA
AGCAGTATGGCGATATTCGGCAACAGCAAAGGTATGGCGGCATTGGCGATGCTGGTGGTG
GGGCTTTTGGTAACGATGGGCATCGGTTCGTCCTTTTCCACTTTGCCGATTATTGCCGCG
ATTTATGTGCCTTTGTGTGTCGGTTTGGGTTTTTCGCCGCTTGCCACCGTCGCCATTGTC
GGCACGGCGGGGGCGTTGGGCGATGCCGGTTCGCCTGCGTCCGATTCCACGCTGGGCCCG
ACGATGGGGCTGAACGCCGACGGGCAGCACGACCACATCCGCGATTCCGTTATCCCGACC
TTCATCCACTACAACATCCCGCTGCTGATTGCCGGCTGGATTGCCGCGATGGTGCTGTAA
ATGGACGCGGTTCAAGAGTTGGAACGCCGTATTGTCGAACTGGAAATCCAATCCGCGCTT
CAGGAGGACGTAATCGCCGGCCTGAACGCGATGGTGGCGGAATTGCGGCAGACGCTGGAT
TTGCAGCAGGCTCAGTTGAGGCTGCTGTATCAAAAAATGCAGGACAGGAATCCCGACGCG
CAAGAGCCGTATTCCCTGCGCGACGAGATTCCGCCGCATTATTGATGCGCCGCCGTATCC
GGATTTCCTTAAAAAAGGCTGTGTTTGAATATTCCGCCGATGCAAACCTAAGATATATAG
TGGATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAAATAGTACGGAAC
CGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGCCAA
CGCTGTACTGGTTTTTGTTAATCCACTATAACTTTAATCAGTCGGACAAAATGCCTTTTA
TCCGATGGAACCGTTTTCCGCCCGAACGGAAATAAGCGGCTTCCCCCGCTGCAAACAGAT
AAAACCAACCGGGTATTCAAACACAGCCAAAAAAAAGCCGTCCGAACCCAAAGGGGACAG
ACGGCCAAACACATTACGGGGAAAACGTTTTACTCAATGAGTCTGCGAAACAGACATTGC
TAAAACAATTGCAATTATAAGGTAAGTTTATTTTTTTTTTTTTTTTTTTGCAAAGAAATG
TAAATTTTTAAATTTTTAAAACCGTTCCGCCGCATCCGTTTGGATGCCGTCTGAAACGGC
GGCAGGGTTTATCGCGCTGTGCCGCAAACCGGGCATTCAGGGTTGCGCGGCAGGTCGAAA
TATTGCCAGCCCCCTTCCAAGGCACGGTAAACCGCCAGCCTGCCGTGCGACGGTTCGCCC
GCATCCAGCAGGATTTTCAGAGCCTCCGCCGCTTGGGTACTGCCGATGATGCCGACCAGC
GGCGAGAACACGCCGAAGAGAGAACAGATGCCGTCTGAAGCCGATCCGCCGTCAAACAGG
CAGGCGTAACACGGCGAGTCGGGCAAGTCGGGACGGTACACGGCAAGTTGCCCTTCAAAG
CGTACCGCCGCCCCTGAAACCAGCGGTGTTTTCGTTTGCACGCAGGCACGGTTGACGGCT
TGCCGCGTGGCGTAGTTGTCGCAACAGTCTAAAACGATGTCGGCGGCTTGAACCAAACCG
GTCAGGCGGCAGCCGTCGAGTTTTTCGTTGACGGCGCGGACGTTGACGGTATGGTTGATG
CGTTTCAGGCGGCCTGCCAAGGCTTCGGTTTTGAGTTTGCCGACATCGCCCTCGTCAAAT
GCGACTTGGCGTTGCAGGTTGTGCAGTTCGACCGTGTCGGAATCGGCTATGGTCAGCGTG
CCGACACCCGAAGCGGCAAGGTAGGGCAGTGCGGCGGCACCCAAACCGCCGCAGCCGACG
ACCAAAATATGCGCGGCGGAAAGTTTCTGCTGCCCTTCGATGCCGATTTCGTCCAAGAGG
ATGTGGCGGCTGTACCGCAGCAGGAATGCATCGTCGTTGTCGTGTTCGGTCGTGGTCATG
ATGATGTTCGGAAAAAAACAGTTGCGGGCGATTGTAACGCTGCCGTCGGGCGGCGTTCAA
CTTCAGACGGCATTTCGGGACACGGGCGGTTAAAGTGTGAACGGTTTGGCACGGATGCGG
CATTTGGGGTACATTTACAATATTTACGGCAGACGAGAGAGAAAAATCATGCAACTGCAT
ATTCTGAACAATCCAAAGGACGCGGCTTTGGCGGCGGACGCGGAATTTCTGAAACAATCC
CTGTTCAACCTCCTGCACGAAGAAGCCTCGCCGTTGGTTGTCGAAACAGTCAAACTCTTG
TCCACTTCCGACGACAGCGCGGCATTGATTGAAAAAGTATTGCCGCAATTGGACGAACAA
CAAACCCACGATTTAACCTTGGCCTGCGGCCTGTTCGCCCAGATTTTGAACATCGCCGAA
GACGTGCACCACGAACGCCGCCGCCAAATCCACGAAGAAGCCGGACGCGGCGGCGCGGAA
GGCAGCCTGACGGAAACCGTCCGCAGGCTCAAAGCGGGGAAAGCCGACGGCAAATCGGTG
CAGCGGCAGTTGGACAATACGTCCGTTACCGCCGTTTTGACCGCGCACCCGACCGAAGTG
CAACGCCAAACCGTCTTAAGCTTCAACCGCCGCATCCGCGCACTGTTGCCGCAACGCGAA
CGCTGCACCAATGCCGACGCGCTGGCACGGCTGCGCCGCGAAATCGACACTATCCTGCTG
GGCTTGTGGCAGACCAGCGAAACGCGCCGCCACAAACTCAGCGTCAACGACGAAATCAAC
AACGGCGTGTCCATCTTCCCGATGAGCTTTTTCGAAGCCCTGCCCAAGCTCTACCGCAAG
ATGGAACACGACTTTCAGACGGCCTATCCCGGCGTCCGCGTTCCGGACATCCTCAAAATC
GGCGGCTGGATCGGCGGCGACCGCGACGGCAATCCGTTTGTTTCTGCCGAAACCCTGCGC
TTTGCCTTCCGCCGCCACGCCGATGCCGTGTTCCGCTTCTATCGCGGCGAACTCGACAAA

CTCTACCGCGAACTGCCGCTCTCCATCCGCCGCGTCAAAGTCAACGGCGATGTAACGGCG
TTGTCCGACAAATCGCCCGACGAAGAAATCGCCCGCGCCGAAGAACCCTACCGCCGCGCC
ATCGCCTACATTATGGCGCGCGCTATGGGCAAAGCGCGCGCGCTCGGTTTGGGTATGGGC
TGCAAATTCGGCTTTCTCGAGCCTTATGCTTCGGCACAAGAGTTTCTGGATGATTTGAAA
AAATTGCAACGTTCCCTTATCGACAACGGCAGCCGTCTGCTTGCCGAAGGCCGTTTGGCA
GACCTCATCCGTTCCGTATCCGTGTTCGGCTTTCACATGATGCCGCTCGACTTGCGCCAA
CACGCAGGCAAACACGCCGATGTGGTTGCCGAGCTTTTCCAACACGCAGGCTTGGAAGAC
TACAACCGCCTGAACGAAGAGCAAAAACAAACCGCCCTGTTGCGCGAATTGAGCCATCAA
CGTCCTCTGTACAGCCCGTTTATCACATACAGCGACCATACCCGCCACGAACTGGCAATT
TTCAACGAAGCGCGCAAAATCAAAGACGAATTTGGCGAAGATGCCGTAACACAAAGCATT
ATTTCCAACTGCGAACAACCCAGCGACCTGCTCGCCTTGGCATTGCTGCTGAAAGAAACC
GGCCTGTTGGCGGTGGAAAACGGCAAACCGCACAGCCGCATCAATATCGTGCCGCTGTTT
GAAACCATTGAAGCGTTGGAAAACGCCTGTCCGGTCATGGAAACCATGTTCCGCCTCGAC
TGGTACGATGCACTGCTCGAAAGCCGTGGAAACATCCAAGAAATCATGCTCGGCTATTCC
GACTCCAACAAGGACGGCGGCTACGTTACCAGCTCATGGTGCCTCTATCAGGCGGAATTG
GGCTTGGTCGAACTCTTCAAAAAATACGATGTCCGTATGCGCCTGTTCCACGGACGCGGC
GGCAGCGTAGGTCGCGGCGGCCCTTCTTACCAAGCCATTCTCGCCCAACCGGCGGGC
AGCGTGGCGGGACAAATCCGCATCACCGAACAAGGCGAAGTCATTACCGCCAAATACGCC
GACCCCGGCAATGCCCAACGCAACTTGGAAACCTTGGTTGCCGCGACTTTGGAAGCCAGC
ATCCTGCCGGATAAAAAAGACCCTGATGCCAAACTGATGCAGGCATTGTCGGACGTATCG
TTCAAATACTACCGCGAACTGATTACCCATCCCGACTTCATCGACTACTTTCTGCAAACC
AGCCCGATTCAGGAAATCGCCACCCTCAACCTAGGCAGCCGTCCCGCCAGCCGCAAAACC
TTGGCGCGGATTCAGGACTTGCGCGCGATTCCGTGGGTATTTTCCTGGATGCAGAACCGC
CTCATGCTGCCGGCTTGGTACGGTTTCGGCAGCGCGGTGGAAACCTTGTGCGAAGACAAA
CCCGAAACGCTCGCCGCCCTGCGCGAACACGCCCAAAGCAACCCGTTCTTCCAAGCCATG
CTCTCCAATATGGAACAAGTGATGGCGAAAACCGACATCACCCTCGCGGAAAACTATGCC
GGCTTGAGCGAATCGCCCGATAAGGCAAAAATCATCTTCGGGATGATTAAGGAAGAATAC
CGCCGCAGCCGCAAAGCACTGCTCGACCTACTGCAAACCGAAGAGCTTTTGCGCGACAAC
CGCAGCCTCGCCCGTTCGCTCGCTTTGAGGATTCCCTACCTGAACGCGCTCAACGGTTTG
CAAGTCGCCATGCTCAAACGCCTGCGTAAAGAACCCGACAATCCGCACGCCCTTCTGATG
GTGCACCTGACCATCAACGGCGTGGCGCAAGGTTTGCGCAATACAGGCTGATAGTGCCGC
ATCGGGGCAAAATGCCGTCTGAACGCCTTTCAGACGGCATTTCCCTGACCGCACTTGCAG
AGAAACACCGATTGTTTTAAAGTGAACGGCAGTGATATGTTGAAAGACGACCAATGAAAA
TTACCGTTATCGGCGCAGGTTCGTGGGGTACGGCGCTCGCCCTGCATTTTTCCCAACACG
GCAACCGCGTATCCCTGTGGACGCGCAACGCAGACCAAGTCCGTCAAATGCAGGAAGCGC
GTGAAAACAAACGCGGACTGCCCGGCTTTTCCTTTCCCGAAACCTTGGAAGTGTGTGCGG
ATTTGGCAGACGCGCTCAAAGACAGCGGACTTGTCCTTATCGTAACCTCCGTTGCCGGAT
TGAGAAGCAGCGCAGAGCTGCTCAAACAGTACGGCGCGGGACACCTCCCCGTCCTCGCCG
CCTGCAAAGGATTCGAGCAGGATACCGGGCTGCTGACCTTTCAAGTCTTGAAAGAAGTAT
TGCCCGACAATAAGAAAATCGGCGTACTTTCCGGCCCGAGTTTTGCACAGGAACTCGCCA
AACAACTGCCCTGCGCCGTCGTCCTTGCCTCCGAAAACCAAGAGTGGATTGAAGAACTCG
TACCGCAGCTCAACACGACCGTCATGAGGCTTTACGGCAGTACCGATGTTATCGGCGTGG
CGGTTGGCGGCGCGGTAAAAAATGTTATGGCGATTGCCACCGGATTGTCCGACGGCCTAG
AGTACGGGCTTAACGCCCGTGCCGCACTGGTTACGCGCGGATTAGCTGAAATCACCCGCC
TTGCCTCCGCAATGGGCGCACAGCCCAAAACCATGATGGGGCTGGCAGGCATCGGCGACC
TCATCCTCACCTGCACCGGCGCACTTTCGCGCAACCGCCGCGTCGGCTTGGGTTTGGCAG
AAGGCAAGGAACTGCATCAGGTGCTGGTCGAAATCGGACACGTTTCCGAAGGGGTCAGCA
CGATAGAAGAAGTCTTCAATACTGCCTGTAAGTACCAAATCGACATGCCGATTACCCAAA
CTCTGCTGCAACTCATCCGCAAAGAAATGACCCCGCAACAGGTTGTCGAAAGACTGATGG
AACGCAGCGCGCGTTTTGAATAAACAACAGACAGATGCCGTCTGAAGCCTTCAGACGGCA
TACGGACAGGTAAGGTTATGAAACAAAATATCGAAAAACTCGAAAGCAGCGTTTATACGT
TGGTACAAAAATTCGAAACCCTCGTCAGCGAAAACCGCCGCCTCAAAGAAACCGTCGCCG
AACTCAAACGGGCGCACGAGCGGCAAAAACTCGAACACGAAACCGCCGTCGACGAACTCA
GCGAAGCCCTGCTCGTCCAAGTCGGCAAACTCAAAGAAGACCTGCAAAACAAAATTGACA
GCCTGACAGAAGAAAATACACGATACCGCAGCCTGCTCGAACAGAGCAGGGAAAAAATCA
GCGCACTGGCAGCGCGCCTCCCCCAATGGCAGGAAACGCAGCAATAAGGATTAAAGGATG
AACATCGAACAAGTCCACATCGAAGTCATGCACGCCCGGCTGACCGTCAACACGCCGGCA

GAAGAAAAAGACACACTGTTGCAGGCAGTCGGAATGCTCAACGGCAAAGCCGAAGCCATC
CGCGAAGGCGGACGCGTCGCGGACAGCGAAAAAATCGTCATTATGGCCGCGCTCAACGTC
GTCCACGACCTGTTGAAAACCTCCCTGAACGGCGGCGATTTGGCAATCGGCGATTTTGCG
CGTAAAATAGCCGATATGGACAACGCCTGCCAAAAAGCACTATCCCGCTTGGCGCAGGAA
TAACCCCTTTTTCCCTGCGGTGTCCGCGAAGGCATATATTCCTTTGAACCAATAAGTTCG
CTTAAGGTTGCCGGGAGCAGTAGTGGGCGCGAGCGTCCTTTTGCGGACGCACCCGAAACT
ACCCGAGGCAGCCGCCTTGTCAGCAAGGTTCAAGCGGATTCGGCCGAAACGGCTCTCGCG
GGGATTCCCATTCAAAACCGCATCGCAACGATGCGGTTTTTCACATTGGCCGCCGCCCCG
CCCGAAAAAGCAGTATGCAGTCTGCCGAAATCCGGAAAATCCTTCCGTAACGCGGTAATG
CGATTGACAAAATCGAATCCTGTCTTTAAAATTTACAACTTTCTATATCTATCTGGATTT
CCACTATGAAAACCTTTTCAGCGAAACCCCACGAGGTGAAGCGCGAATGGTTCGTCATCG
ATGCCCAAGACAAAGTCTTGGGTCGCGTTGCGGCCGAAGTCGCCAGCCGTCTGCGTGGCA
AACACAAACCTGAATACACCCCCCACGTCGATACCGGCGATTACATCATTGTTATCAATG
CGGACAAACTGCGTGTAACCGGTGCCAAATTCGAAGATAAAAAATACTTCCGCCATTCCG
GTTTCCCAGGCGGTATCTACGAACGCACCTTCCGCGAAATGCAAGAGCAATTCCCGGGCC
GCGCTTTGGAACAAGCTGTAAAAGGTATGCTGCCCAAAGGTCCTCTGGGTTACGCCATGA
TTAAAAAACTGAAAGTGTATGCGGGTGCGGAACACGCCCATGCTGCGCAACAACCCAAAG
TTTTGGAACTGAAATAAGGACGCGACATGAACGGTAAATACTACTACGGCACAGGCCGCC
GCAAAAGTTCAGTGGCTCGTGTATTCCTGATTAAAGGTACAGGTCAAATCATCGTAAACG
GTCGTCCCGTTGACGAATTCTTCGCACGGGAAACCAGCCGAATGGTTGTTCGCCAACCCT
TGGTTCTGACTGAAAACGCCGAATCTTTCGACATCAAAGTCAATGTTGTTGGCGGCGGCG
AAACCGGCCAGTCCGGCGCAATCCGCCACGGCATTACCCGTGCCCTGATCGACTTCGATG
CCGCGTTGAAACCCGCCTTGTCTCAAGCTGGTTTTGTTACCCGCGATGCCCGCGAAGTCG
AACGTAAAAAACCGGGTCTGCGCAAAGCACGCCGTGCAAAACAATTCTCCAAACGTTAAT
GTTGGAAATTCAAAAAACCCTGCTTATCGCAGGGTTTTTATTTGTAGTAGACGGTTTCC
CATTGGCAATCTAAAGATTACAGATTGGGCAAAAATCAAAAACAGTATAGTGGATTAACA
AAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAAATAGTATGGCAAGGCGAGGCAA
CGCCGTACTGGTTTTTGTTAATCCACTATACATGAAAAAATAGAAAAACTGCCGGGTATG
GTTGAATAAGGGGTCAGACCGGTTCCAGTTCGCTCAGTCCGGGCAAATCTGCAAAACCGC
GTTCGCGTATGATTTGGCAAAAATTGTTCAGATAGCTCTTGTCCGTATCTTCGGTACGGA
TGGCGGCATACAGTTTGCTTTGCAGTCCGTCGGCAGTAATTTGGCGGTGGACGACATAGC
CTTTTTCAAGGTAGGGCATGACTGTCCAATAGGGAAGGGCGGCAATGCCACGTCTGCTGG
CAACCAGTTGGATAATGGCGATGGTCAGCTCGCTGTGTCGGCGCGGCGGGTTGATGTTTT
TCGGAATCAGGATTTTTTTGGGCAAATCCAGCATCTCGTCGGGAACGGGATAAGTAATCA
GGGTTTCCCCGATAAAGTCTTCCGCCGTCCAAACGTTTTTGGCGGCAAGCGGATGGTCTG
GTGCGCAAATGCCGACCATTTCGTAGGCAAACAGCGGTTGGAAAGAAATACCGTTTTGTT
TTTCCGCTTCGGAAACAATGGCAAGGTCGGCACGGTGTTGCAGCAGCAGTCCGACGGGAT
CCGCTTGGAATCCCGATACGATATCCAATTCGACTTGGGGCCACATCGGGCGGAATTCGC
CCATGGCGGGCATCAGCCAGTCGAAACAGGTATGGCATTCGACGGCAATCCGCAGCTCTC
CCGCCTCTCCTTCCGTGATTCGCGCCAAATCCCATTCTGCAACAGCAACTTGAGGTATAA
GTTCGTGGGCGAGGCGCAGCAGCCTTTCGCCCACCGGGGTAAAGCGCAAGGGCGTGGATT
TGCGTTCGAACAGCGGCGTGCCGTAGTGGTTTTCGAGCATACGGATCTGGTGGGAAAGGG
CGGATTGGGTAAGGAAAACCCGTTTGGCGGCAAGGGAGACGCTGCCGGTTTCTTCAAGTG
CCAGCAGGGTTTTGAGGTGGCGCAATTCGATAATGGAATCCATGGGGCTTCAGACGGCAT
ATTGAACCGGCGCATATTAAAATAATTCATATGCGGGTGCAAATCGGAAAAATGTGAATT
CCGGACATTTTGCGATTAGAATGCCCGCTTGTTTAAAGCGATTAGGAAAAAAGATGGTAT
TGTGCAGGGATTTTCTGACTTGGTGTAATGAAACATTGCAGACAGCGTTGTTTAAAGATT
ACGCCCCTAACGGTTTGCAGGTTGAAGGGAGGGAATATATCGGGAAAATCGTTACGTCGG
TAACGGCAAGCAGGGCAGCGATTGATTTTGCTGTGGAGCAGAAGGCAGATTTGCTTTTGG
TACATCACGGTATGTTCTGGAAAAACGAGTTGCCGACCGTTACTGGTTGGAAAAAAGAAC
GGATTGCCGCACTGTTACGGCACGACATCAATATGGCAGGCTACCATCTGCCCCTGGATG
CACATCCCACACTGGGCAACAATGCCCAACTCGCCGACAGATTGGGTTTTGCGACAGAAA
AACGGTTCGGCGAACAAAACCTGCTCAACTCGGGCAGCCTGAAACAAGCCAAGACACTCG
GCGCATTGGCGGCGCATATTGAAACAGTTTTGCAACGTAAGCCTGTCGTTATCGGCAATC
CCGAACGCGAAATCCGACGGGTTGCATGGTGCAGCGGCGGGGCGCAGGGGTTTTTTCAGA
CGGCAATAGACGAAGGTGTCGATCTGTATTTGACGGGGGAAATCTCTGAAGCCCAATACC
ACCTTGCCAATGAAACGGGTACGGCTTTCATTTCGGCAGGGCATCACGCGACGGAACGTT

ACGGCGTACGCGCGCTGGCAGAATCGGCGGCAGAGGTTTTCGGGTTGGAAGTGTGCCATT
TTGACGAAAACAATCCGGCTTGAGTCTTGAGAAATATCATAAAACTTTACCTTATTTTAA
ATAATGCTTTGAGTATCACTGCAAACTGGGTAAAATTGCAACGTTTAATGGCTCGGTATT
CCCAAAATATTAGGACGACTGAATAATGGATAATCAAGAAATCAACAACGGCCGCCGCCG
TTTCCTGACACTCGCGACCTGCGGCGCGGGCGGAGTGGCAGCATTGGGTGTGGCAACGCC
GTTTGTGGCCAGTTTTTCCCTTCGGAAAAAGCCAAGGCCGCCGGTGCTGCCGTCGAGGT
GGATGTCAGTAAAATCGAAGCGGGTCAGCTGCTGACCGCCGAGTGGCAAGGCAAACCGAT
TTGGGTGCTCAACCGTACAGATCAGCAGCTTAAAGACCTGAAAGGCCTGAACGGCGAACT
TACCGATCCCAATTCCGATGCGGAACAGCAGCCGGAGTATGCTAAAAACGAGACCCGTTC
GATTAAGCCGAACATCCTTGTCGCCATCGGTATCTGCACCCATTTGGGCTGCTCGCCCAC
CTTCCGTCCCGACATTGCCCCCGCCGATTTGGGTGCAGACTGGAAAGGCGGCTTCTTCTG
CCCCTGCCACGGTTCGAAATTCGACTTGGCCGGCCGCGTATATAAAGGTGTTCCTGCCCC
GACCAACCTGGTTGTCCCGCCATATAAATACTTGAGCGACACAACTATCTTGGTGGGCGA
AGACTAAGAATAAGGAACGATAATTATGGCAAACCAAACCAATAGCAAAGCAAAAGCATT
GTTAGGCTGGGTAGATGCCCGTTTTCCATTAAGTAAAATGTGGAAAGAGCATCTGTCTGA
ATACTATGCGCCTAAAAACTTCAACTTCTGGTATTTCTTCGGCTCATTGTCTATGCTGGT
GCTGGTGATTCAAATCGTCAGCGGTATTTTCCTGACCATGAACTACAAACCGGACGGCAA
CCTTAACGCCTACCATCTGCCTGCTGCCTTTACCGCAGTAGAGTACATCATGCGCGACGT
GTCCGGCGGCTGGATTATCCGCTATATGCACTCTACCGGCGCATCTTTCTTCTTCATCGT
CGTTTATCTGCACATGTTCCGTGGTCTGATTTACGGTTCGTACAAAAAACCGCGCGAATT
GGTGTGGATTTTCGGTTCCCTGATTTTCTTGGCATTGATGGCAGAAGCCTTTATGGGCTA
CCTGCTGCCTTGGGGTCAAATGTCCTTCTGGGGTGCGCAGGTAATTATTAACCTGTTCTC
CGCCATCCCTGTTATCGGTCCTGATTTGTCCACTTGGATCCGCGGTGACTTCAACGTTTC
CGATGTTACTTTGAACCGATTCTTCGCCCTGCACGTTATCGCTGTACCTTTGGTATTGCT
CGGCTTGGTTGTGGCTCATATCATTGCCTTGCATGAAGTGGGTTCCAACAACCCTGACGG
TGTAGAAATCAAAAAGCTGAAAGATGAAAACGGTGTCCCTCTAGATGGCATACCTTTTTT
TCCGTATTATGTTGTGCATGATATATTGGCAGTAACGATATTCTTGATTGTCTTCTGTGC
CGTGATGTTCTTTGCACCTGAAGGCGGCGGCTACTTCTTGGAAGCGCCAAACTTCGATGC
AGCGAATGCGCTGAAAACACCTCCGCACATTGCGCCGGTATGGTACTTCACTCCGTTCTA
CGCAATTCTGCGTGCGATTCCTTCCTTTGCCGGTACTCAGGTATGGGGTGTAATCGGTAT
GGGTGCAGCAGTTGTACTGATCGCCTTGCTGCCTTGGTTGGATAAAGGCGAGGTTAAATC
TGTCCGCTATCGCGGCCCAATCTTCAAAACCGCATTGGTTCTGTTCATCATTGCCTTCAT
CGGTTTGGGTATTTTGGGTGCAATGGTAGCAACTGATACGCGTACTTTGGTTGCACGTAT
CCTGTCTTTCGTCTACTTTGCATTCTTCCTGGGTATGCCGTTCTATACCAAACTGGATAC
CAACAAACCAGTTCCTGAACGCGTAACCATGAGCACTACTAAACAAAAAATTATGTTCTT
TGTTTACGTCGGTATTACCGTTGTTGGTGCTTACTTGTTTGCAACCAATATCTGATGAGG
GCAGCGAAAATGAAACAAACTCTGAAAAACTGGTTTGCTGCCTTATTGCTGGCAGTGCCT
ATGAGTGCAGCCGTCGCCAGCGGCGGCGGACACTACGAAAAAGTCGATATCGACCTGCGT
GACCAAGTCAGCCTGCAGCACGGTGCGCAAATCTTTACAAACTACTGTTTGTCTTGCCAC
TCTGCAAGCGGTATGCGCTTCAACCGTCTGAAAGACATCGGTTTGACTGACGAAGAAATC
AAGAAAAACCTGATGTTTACCACCGATAATGTCGGCGATGTCATGCATTCGGCGATGAAC
CCGAAAGATGCGGCAAAATGGTTTGGTGCTGCTCCGCCCGATTTGACGTTGATTGCGCGT
TCCAAAGGTGCAGACTACCTTTACGCCTATATGCGCGGCTTCTATAAAGATCCGACCCGT
CCGAGCGGCTGGAACAATACTGTATTCGATAAAGTCGGTATGCCCCACCCGTTGTGGGAG
CAACAAGGTGTTCAAGCCGTTGAGTTGGATGCCAAAGGTCAGCCGGTTATGGTAAAAGAC
GAACACGGCGAGATGAAGCCTAAGCTGTATTGGGAATCTACCGGTTTGCACAGCCGCCGC
CTGCCTAACGGCAAAGTGATCCAAAAAGAGTACGACGCATATGTACGCGATTTGGTCAAT
TACCTTGTGTACATGGGCGAACCTGCACAACTGCAACGCAAACGTATAGGCTATGTCGTG
ATGATTTTCCTATTTGCGGTTATGCTGCCTTTGGCTTACTTCCTAAATAAAGAATATTGG
AAAGACGTACACTAAGCGTTTGGAACAAAAGGGCAAATCCTTTAGGGTTTGCCCTTTTTT
CATTTTGCCTGCCGTTTGAAAAGCCTGAATCCGTATGCCGTCTGAACATAGCTGCAACAT
TTCAGACGGCATCATTCTAAAAATGTCAGACATCGGGAACTTAATCAGGGTTTGTGGCAG
ATTGATATTGAATTAGGAAAAGGTCTGAAACCGCAAGCAGCATGGTAAGGCGATGTCGAA
CAGGTTTAAAGCCAATCCGCTATATTTTTTGGAGCTATGTCGAACAAACAGGTAGAATCG
GAATAAGCGGCGGTTTTGTCCGGTGTGAAAATGTTTGAATATTAGGATAATTGATGGAAC
TGATGACTGTTTTGCTGCCTTTGGCGGCGTTGGTGTCGGGCGTGTTGTTTACATGGTTGC
TGATGAAGGGCCGGTTTCAGGGCGAGTTTGCCGGTTTGAACGCGCACCTGGCGGAAAAGG

CGGCAAGATGTGATTTTGTCGAACAGGCACACGGCAAAACCGTGTCGGAATTGGCGGTGT
TGGACGGGAAATACCGGCATTTGCAGGACGAAAATTATGCTTTGGGCAACCGTTTTTCCG
CAGCCGAAAAGCAGATTGCCCATTTGCAGGAAAAAGAGGCGGAGTCGGCGCGGCTGAAGC
AGTCGTATATCGAGTTGCAGGAAAAGGCACAGGGTTTGGCGGTTGAAAACGAACGTTTGG
CAACGCAGCTCGGACAGGAACGGAAGGCGTTTGCCGACCAATATGCCTTGGAACGCCAAA
TCCGCCAAAGAATCGAAACCGATTTGGAAGAAAGCCGCCAAACTGTCCGCGACGTGCAAA
ACGACCTTTCCGATGTCGGCAACCGTTTTGCCGCAGCCGAAAAACAGATTGCCCATTTGC
AGGAAAAAGAGGCGGAAGCGGAGCGGTTGAGGCAGTCGCATACCGAGTTGCAGGAAAAGG
CACAGGGTTTGGCGGTTGAAAACGAACGTTTGGCAACGCAAATCGAACAGGAACGCCTTG
CTTCTGAAGAGAAGCTGTCCTTGCTGGGCGAGGCGCGCAAAAGTTTGAGCGATCAGTTTC
AAAATCTTGCCAACACGATTTTGGAAGAAAAAGCCGCCGTTTTACCGAGCAGAACCGCG
AGCAGCTCCATCAGGTTTTGAACCCGCTAAACGAACGCATCCACGGTTTCGGCGAGTTGG
TCAAGCAAACCTATGATAAAGAATCGCGCGAGCGGCTGACGTTGGAAAACGAATTGAAAC
GGCTTCAGGGGTTGAACGCGCAGCTGCACAGCGAGGCAAAGGCCCTGACCAACGCGCTGA
CCGGTACGCAGAATAAGGTTCAGGGCAATTGGGGCGAGATGATTCTGGAAACGGTTTTGG
AAAATTCCGGCCTTCAGAAAGGGCGGGAATATGTGGTTCAGGCGGCATCCGTCCGAAAAG
AGGAAGACGGCGGCACGCGCCGCCTCCAGCCCGACGTTTTGGTCAACCTGCCCGACAACA
AGCAGATTGTGATTGATTCCAAGGTCTCGCTGACAGCTTATGTGCGCTACACGCAGGCGG
CGGATGCGGATACGGCGGCACGCGAACTGGCGGCACACGTTGCCAGCATCCGTGCACACA
TGAAAGGCTTGTCGCTGAAGGATTACACCGATTTGGAAGGTGTGAACACATTGGATTTCG
TCTTTATGTTTATCCCTGTCGAACCGGCCTACCTGTTGGCGTTGCAGAATGACGCGGGCT
TGTTCCAAGAGTGTTTCGACAAACGGATTATGCTGGTCGGCCCCAGTACGCTGCTGGCGA
CTTTGAGGACGGTGGCGAATATTTGGCGCAACGAACAGCAAAATCAGAACGCACTGGCGA
TTGCGGACGAAGGCGGCAAGCTGTACGACAAGTTTGTCGGCTTCGTACAGACGCTCGAAA
GCGTCGGCAAAGGCATCGATCAGGCGCAAAGCAGTTTTCAGACGGCATTCAAGCAACTTG
CCGAAGGGCGCGCGGGAATCTGGTCGGACGCGCCGAGAAACTGCGTCTGTTGGGCGTGAA
GGCAGGCAAACAACTTCAACGGGATTTGGTCGAGCGTTCCAATGAAACAACGGCGTTGTC
GGAATCTTTGGAATACGCGGCAGAAGATGAAGCAGTCTGACTTGTGCGGAAAAATATTGT
TTCAGCCGGGGCGGGAATGCCGAAAGCGCGGCGTAACTGTACTGGTTCTGATTTTGGGTT
TTTTGTTTGAAGTACTTACCAAAGCCTTGTCCGGCGAGGTACACCGGCAAGGCGGCGGTT
GTCTGAGCCTTTTGTGTTTCAGACGGCATCAGTGCAGATGATGCCGTCTGAAGACCGTAG
GGAAGGCGGTTAGAAAAACGGATTGTGCCGCTTTTCGTGTCCGATGGAAGTCATACGCCC
GTGTCCGGCGACAACTTGCACGGTTTCGGGAAGGGTGAATAATTTGTTGCGGATATTATT
GATTAAGTCGGCGTGGTTGCCGCGCGGAAAATCGGTTCTGCCTATGGTTTCGTAAAACAG
CACGTCGCCCGCAATCAGCAATTCCGCCTCGGCACAATAAAAGACGATATGTCCCGGCGT
ATGGCCCGGAATATGCAGCACTTGAAAGGCATAGCGTCCGACCGTGAGCGTTTCGCCTTC
TTCGAGCCAACGGTTCGGCGCAAAGGCGGGCGAGACGGGAAATCCGTATTGCGCGGTGGT
TTGCGGCAGCGATTGGAGCAGGAATTCATCGTCCGGATGCGGCCCGAGGACAGGGACTTT
ATGCGTTTTCAACATTTCGACCACGCCGCCCGCGTGATCGAGATGGCCGTGCGTCAGCCA
GATTGCCGTGAGCGTAAGTTTGCGGTTTGCCAACGCTTGCAGCAGGAACGGCACGTCGCC
GCCGACATCGGTCAGGACGGCTTCGCCGCTTTCGTCGTCCCAAATCAGGGTGCAGTTTTG
GCGGAAGGGGGTAACGGGGAAGATTTCGTAACGTAGGGTCATCGGCGTGTTCCTAAACGG
TTTTTCAGACGGCATCGGGTTTGCCGTTTGTTTTATGGCGGTTTGCCGCCCGTTTTGATA
TTGGGGGAATCAGATGATTAAGAAGACAATCGGCGGCATCATACCGATTTTTACGGCGGT
TTTCATCCCTGCATCAGCAGGCGCGGCGGATTTGATGCTGGCGCAGGAATACAAAGGGCA
GGATATTGCCGGCTGGGCGATGAGCGAGAAACTCGACGGCGTGCGCGCCTATTGGGACGG
AAAGCACCTGATGAGCCGTCAGGGCTACGCGTTTGCTCCGCCCAAAGGTTTTACCGCTCA
GTTTCCGCCTTATCCTTTGGACGGCGAATTGTATAGCGGACGTGGTCAGTTCGAGCAGAT
TTCCGCTACCGTGCGTTCTGTTTCTTCAGACTGGCGCGGCATCCGCCTGCACGTTTTCGA
TGTACCCAAGGCGCAGGGCAACCTCTACCAACGTTTGGCAGTCGCAACGCAGTGGCTGAA
AACGCATCCGAACGCGCCGATTACCATCATCCCGCAAATCAAAGTGCGCGACCGGCAGCA
CGCGATGGACTTTTTAAAACAAATCGAAGCGCAGGGCGGCGAAGGCGTGATGCTGCGTCA
GCCCGAATCCCGTTACAGCGGCGGCAGGAGCAGCCAATTATTGAAGCTGAAAAGCCAATA
CGACGACGAATGCACGGTAACGCGGCACTATGAGGGCAAAGGGCGAAACGCCGGACGGCT
GGGCGCGGTCGGCTGCAAAAACCGGCACGGCGAATTCCGCATCGGCAGCGGTTTCAAAGA
TAAAGACCGCGACAACCCGCCCAAAATCGGCACACTGATTACCTACCGTTACCGTGGCTT
TACGCGGAAAGGCACGCCGAAATTTGCCACATTTGTGCGCGTGCGTACCGACCGCTGAGC

AGGACAGTTCAGACGGCATCCGATACTTTGGTTTATAATTTTCCTTTTACCGACCGATTC
CGACATATGACCGATTTAGAAACCAAACGCCTTGAAACACAGGCGATGCTTGAAAACGCC
GATCTTTTGTTCGACCAAGGCCAATGCCGTGCCGCACTGCAAAAAGTGGCGGACGAGATT
ACGCGTGATTTGGGCGGCAAATATCCGCTGCTGCTGCCCGTGATGGGCGGCGCGGTGGTG
TTTACGGGGCAGTTGCTGCCGCTGTTGCGTTTTCCCTTAGATTTTGATTATGTTCACGTT
TCCCGTTACGGCGACAAGCTGGAGGGCGGCGCGTTCAACTGGAAGCGTATGCCCGATGCG
GAACAAATCCGGGGCAGGCACGTCGTCGTGCTGGACGATATTTTGGACGAAGGGCATACG
ATGTCCGCCATTCAAGCCAAACTTTTGGAAATGGGTGCGGCAAGCTGCCGTGCGGCGGTG
TTCGCCAACAAGCTGATTGACAAACCCAAGCCTATCCGAGCCGATTATGTCGGACTGGAT
GTGCCGAACCGTTATGTTTTCGGTTACGGCATGGATGCGGCGGGCTGCTGGCGCAATCTG
GGCGAGATTTACGCATTGGGCGGAAAATAAGGGCGCGATGCCGTCTGAAGGCTGTTCAGA
CGGCATCGCGGCCATACGCCGGCAGGATAATGAGGAACAGGACGCATAATATGATAGGGC
TTTTAATCATCACACACGAAACCATAGGCGAAGCCTACCGCAAGCTGGCGCATCATTTTT
TTCCGGGCGGACTGCCTGAAAACGTCCGCATACTCGGCGTGCAGCCGACGGAAGACCAAG
ACGACATCAACAACAACGCCATTGCCGCGCTTCAGGAATTTCCCGACAACGACGGCGTGC
TGATTATGACCGATATTTTCGGTGCGACCCCCTGCAATGCCGCCCGCCGCCTCGTGCGCG
AAAACAAATCGGCGATTTTGACCGGGCTGAACGCGCCGATGATGGTTAAGGCCGTCCAAT
ATTCGCCGGCGGCGGAAGACCTTGCCGCCTTTACCGAATGCGTCAGGGAGGCGGCGGTAA
AAGGCATTTTCGCCATCACGTCCGCGCCCGAAGATTTGGTGTGCCGGCGCAGCGGCGATG
CCGTCTGAAGAAGCGGCAGGGCAGGAAAACATTTTAAACGGCGTCTGCTGCCGATATACA
TAACACGGGAATCGAAATGCTCAAACAATCCATCGAAATCATCAACAAACTCGGACTCCA
CGCCCGCGCGTCCAACAAGTTCACCCAAACCGCGTCCCAATTCAAAAGCGAAGTCTGGGT
TACGAAAAACGACAGCCGCGTCAACGGCAAAAGCATTATGGGGCTGATGATGCTCGCCGC
CGCCAAGGGTACGGTCATCGAACTGGAGACGGACGGCGCGGACGAGGCGGAAGCGATGCG
CGCCCTGACCGACTTAATCAACGGCTACTTCGGCGAGGGCGAATAATGAGTATCGTGCTG
CACGGCGTGGCGGCGGGCAAAGGCATTGCCGTCGGTTGCGCCCACCTGATTGCGCGCGGT
ACGGAGGAAGTGCCGCAGTATGATGTTGCGGAGGCGGACACCGATGCCGAAGCCGAACGT
TTCGATGCCGCCGTCAAAGCCACGCGCAAAGAGTTGGAACAGCTCCGCAGCGCGATTCCC
GAAAACGCCCCGACCGAGTTGGGCGCGTTCATCTCGCTACACCTGATGCTCTTGACCGAT
GTTACCTTGTCGCGCGAACCCGTCGATATTTTAAGGGAACAAAAAATCAACGCCGAGTGG
GCATTGAAGCAGCAGAGCGACAAACTCGCCGCCCAATTCGACAATATGGACGATGCCTAT
TTGCGCGAACGCAAGCAGGATATGCTGCAAGTCGTCCGCCGCATCCACAACAACCTGATC
GGGCAGGGCAACGAGTTGGAAGTTGCCGACAACCTGTTTGACGAAACCGTTCTGATTGCA
AACGACCTTTCGCCCGCCGACACGGTTTTGTTTAAAGAGCAGCGCATTGCCGCCTTCGTT
ACCGATGCCGGCGGCCCCACCGGGCATACGGCGATTTTGGGCAGGAGCTTGGACATCCCG
TCCGTCGTCGGGCTGCACAACGCGCGCAAACTGATTACCGAAGGCGAAACGGTCATTGTG
GACGGTATCAACGGCGTGTTGATTATCGCGCCGGATGAGTCGGTGTTGAACGAATACCGC
CGCCGTGCCCGCGAATACCGCAGCCACAAACGCGATTTGAACAAGCTCAAAAAAACCGCC
GCCGCCACCGCCGACGGGGTCTGCATCGAGCTTGTGGGCAATATAGAATCCGCCGAAGAC
GTGAAACCGCTGCACAACCTCGGCGCAGACGGCATCGGGCTGTTCCGCAGCGAGTTTCTT
TACCTGAACCGCGATACGATGCCGTCTGAAGACGAGCAGTACGAAGTGTACAGCGCGATT
GTCAAAAAAATGAAAGGCAAAAGCGTAACGATACGGACAGTCGATTTAGGTGTGGACAAA
AACCCGCGCTGGTTCGGGAAAAACAGCACGCCCAACGGCAGCCTCAACCCCGCGCTGGGC
ATGACCGGCATCCGCCTGTGCCTTGCCGAACCGGTCATGTTCCGCACCCAGATGCGCGCC
ATCCTCCGTGCGGCGGTACACGGCCCCGTGCGGATGATGTGGCCGATGATTACCTCCGTA
TCCGAAGTGCGCCAGTGCCTCATCCACCTCGACACCGCGCAACGCCAGCTTGCCGAACGC
GGCGATGCCTTCGGTAAAGTCGGCATCGGCTGTATGATTGAAATTCCGTCTGCCGCGCTG
ACCGTCGGCAGTATTTTGAAACTGGTCGATTTCATCTCCGTCGGTACCAACGACCTGATT
CAATACATCTTGTCCGTCGATCGCGGCGACGACAGCGTCAGCCACCTCTACCAGCCCGGC
CATCCCGCCGTGCTGAAAATGCTGCAACACGTCATCCGTACCGCCAACCGCATGGACAAA
GACGTATCCGTATGCGGCGAGATGGCGGGCGATACCGCGTTTACCCGCGTTTTATTGGGT
ATGGGGCTGCGCCGTTTTTCCATGAACCCCAACAACATCCTGCCCGTCAAAAACATCATT
CTGCACAGCAATGTCGGACAGCTCGAAAGTGATATTGTGAAAGTCATCCGCTGCGAAGAC
GAAGAGAAGTCGGAAAAGCTGATCAAACAGATGAACAGCGTGTCTGTCGAGGAAGAAGCC
GACTTCAAGGGGCGGAAATAAATACGGCAGGTAAAAAATAGAAATACTTAACAATGCCCG
CAATCTGAAATTTTGCCATTCTTGCAAAATAGAAAACCGAAACAGAAACCCAAAATCGGC
CATTCCCTCAAAAACAGAAAACCAAAATCAGAAACCTAAAATCCGTCATTCCCGCGCAGG

CGGGAATCTAGGTTTGTCGGCACGGAAACTTATCGGGAAAAACGGTTTCTTTAGATTTTA
CGTTCTAGATTCCCGCCTGCGCGGGAATGACGATGAAAAGATTGTTGTCGCTTCGGATAA
ATTTTTGCCGTGTTGGGTTCTAGATTCCCGCTTTCGCGGGAATGACGGCAGAGTGGTTTC
AGTTGCTCTCGATAAATGCCGCCATCTCAAGTCTCGTCATTCCCTTAAAACAGAAAACCG
AAATCAGAAACCTAAAATCCCGTCATTCCCGCGCAGGCGGGAATCTAGGTCTGTCGGCAC
AGAAACTTGTCGGGAAAAACGGTTTCTTTAGATTTTACGTTCTAGATTCCCGCCTGCGCG
GGAATGACGATGAAAAGATTGTTGTCGCTTCGGATAAATTTTTGTCGCGTTGGGTTCTAG
ATTCCCGCTTTCGCGGGAATGACGGCAGAGTGGTTTCTGTTGCTCCCGATAAATGCCGCC
ATCTCAAGTCTCGTCATTCCCTTAAAACAGAAAACCGAAATCAGAAACCTAAAATCCCGT
CATTCCCGCGCAGGCGGGAATCTAGGTTTGTCGGTGCGGAAACTTGTTGAAAACTTTGCA
AAATCCCCTAAATTCCCACCAAGACATTTAGGAGATTTTCCATGAGCACCTTCTTCCAGC
AAACCGCACAAGCCATGATTGCCAAACACATCAACCGCTTCCCGCTATTGAAGTTGGATC
AAGTGATTGATTGGCAGCCGATCGAACAATACCTGAACCGTCAAAAAACCCGTTACCTTC
GAGACCACCGCGGCCGTCCCGCCTATCCGCTGCTGTCCATGTTCAAAGCCGTCCTGCTCG
GACAATGGCACAGCCTCTCCGATCCCGAACTCGAACACAGCCTCATTACCCGCATCGACT
TCAACCTGTTTTGCCGTTTCGACGAACTGAGCATCCCCGATTACAGCACCTTATGCCGCT
ACCGCAACTGGCTGGCGCAAGACGACACCCTGTCCGAATTGCTCAAACTGATCAACCGCC
AACTGACCGAAAAAGGTTTAAAAGTAGAGAAAGCATCCGCCGCCGTCATTGACGCCACCA
TTATTCAGACCGCCGACGGCAAACAGCGTCAGGCCATAGAAGTCGATGAAGAAGGACAAG
TCAGCGGCCAAACCACACGGAGTAAGGACAGCGATGCGCGTTGGATCAAGAAAAACGGCC
TCTACAAACTCGGTTACAAACAACATACCCGCACCGATGCGGAAGGCTATATCGAGAAAC
TGCACATTATAGTAGATTAAATTTAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACA
GATAGTACGGCAAGGCGAGGCAACGCCGTACTGGTTTAAATTTAATCCACTATATTATGC
GCAAAGCCTGCCGCAACCGTCCGCTGACGGAGGCGCAAACCAAACGCAACCGATATTTGT
CGAAGACCCGTTATGTGGTCGAACAGAGCTTCGGTACGCTGCACCGTAAATTCCGCTACG
CCCGGGCAGCCTATTTCGGACTGATTAAAGTGAGTGCGCAAAGCCATCTGAAGGCGATGT
GTTTGAACCTGTTGAAAGCCGCCAACAGGCTAAGTGCGCCCGCTGCCGCCTAAAAGGCGA
CCGGATGCCTGATTATCGGGTATCCGGGGAGGATTAAGGGGGTATTTGGGTAGAATTAGG
AGGTATTTGGGGCGAAAATAGACGAAAACCTGTGTTTGGGTTTCGGCTGTTGTGAGGGAA
AGGAATTTTGCAAAGGTCTCAGATTGTTGTCGCTTCGGATAAATTTTTGCCGCGTTGGGT
TCTAGATTCCCGCTTTTGCGGGAATGACGGCAGGGTGGTTTCAGTTGCTCCCGATAAATG
CCGCCATCTCAAGTCTCGTCATTCCCTCAAAAACAGAAAACCAAAATCAGAAACCTAAAA
TCCCGTCATTCCCGCGCAGGCGGGAATCTAGGTCTGTCGGCACAGAAACTTGTCGGGAAA
AACGGTTTCTTTAGATTTTACGTTCTGGATTCCCGCCTGCGCGGGAATGACGATGAAAAG
ATTGTTGTCGCTTCGGATAAATTTTTGCCGTGTTGGGTTCTAGATTCCCGCTTTTGCGGG
AATGACGGCAGGGTGGTTTCAGTTGCTCCCGATAAATGCCGCCATCTCAAGTCTCGTCAT
TCCCTCAAAAACAGAAAACCAAAATCAGAAACCTAAAATCCGTCATTCCCGCGCAGGCGG
GAATCTAGGTCTGTCGGCACGGAAACTTATCGGGAAAAACGGTTTCTTTAGATTTTACGT
TCTAGATTCCCGCCTGCGCGGGAATGACGATGAAAAGATTGTTGTTGTTTCGGATAAAAT
TTTGCAGCCCTGATAAAAAAATATGGCTGCTTTGGTAAAAAAATGCCGTCTGAAAGGTTT
TCAGACGGCATTTTGTTTTTAAGAAGCATCAGCGGAAGCGGACGATTTCCCGTTCTTCGA
TATGGATGCGTACCGTATCCTTGCCCGATACCGCCCCGGCGTGCCGCATATCGAGGTTCA
GCCACAGGATGCCGTGTTCCGGATGGAGGACGGACAGGCTGAACGATTCGGGCAAACAGG
TACGGGATAATACGCGGCACTCCATGCCGTCTTGGTCGAAACGCACCGCATGTTGCGGAA
TATGGCGGTTATCGTCGGTATTGGGCAAACCCATCAGTCGGGCGACCTGCACGCAGGATG
GTGTTTTGACCAATGTTTCGGGCGTACCGTATTGTAGAATCCTCCCTTTATGCATCACGG
CGATTTCGTCTGCCGTCGTACAGGCTTCTTCGGGCGAATGCGTTACCAAAACGGCAGGGA
TGCCGCCGTTTCGGATACGTTCGGCAGTCATACGGCGCAGCGTGCCGCGCAAATGCGTGT
CCAAACTGGAAAACGATTCGTCCAACAGCAGCAGGGAAGGGCGGACAACCAAAGCGCGCG
CCAACGCCAGCCGTTGCTTCTCGCCTCCGGAAAGTTTTTCAGGCTTGCGGTGCGCCTCGT
TTTCCAGTCCGACTTCGGCAAGTGCCGCCATGGCGAGGCGTTCGGCTTCGGCTTTCGGCA
TTTTTTGCATTTTCAAACCGAATGCCGCATTTTCCAGCGCACTCATATGGGGAAACAGCG
CGTAATCTTGAAACATCAGCGAGATACGGCGTTTTTCGGGCGGCATACGGGTAATGTTTT
CTCCGTTCAGCCATATTTCCCCGCCGTCCGGCCGGACAATCCCCGCAATTATATTCAGCA
GGGTGGATTTTCCGCAGCCCGACCGCCCCAAAACGGCGAGTATTTTGCCGCGCCCGACAG
TCAGGCAGATGTTGTCGGCGACGGTTTTATTGCCGAAGCGTTTGCAGAGTCCGTTCAGTT
CAAGCATGGCGCATCCTATAAACGTATGCCGTCCAGCCACTCGGACAGCGGATGGGCGGC

ATCAATCAGACCGTAACGGCAAAACGCGTCCAGCGTAACCAGTTGCGCGTCGTGCATCAT
GTTTCCCGACAACATGGCATCCAACAGACCGCCGATGTCCATTTTCTCAAAACCCGCCAC
TTCGCCATCCTGATTTTCAGGCAGGAAGGTTTCGGGCAGGACGGCATCGAATACATACAG
GATTTCATTGTGTACACCCCGGCTGACGGAGCGCAGGCTGTGCAGCTGCGATACCGGGCG
GATGAGCGGAAGCAGCGTTTTATCCAAACCGGCTTCTTCGCTGCTTTCGCGACACACGGC
TTCAGACGGCATTTCGCCGCCGGAAACACCGCCGGCGGCAGTATTGTCGAGTTTGTTGGG
ATCGACTGCTTTGTGCGGACTGCGCCTGCCTATCCAGAAATGCCATCGGCCGTCCGATTC
GGTCAGACCGTTGAGATGGACGGCGCGGCTGAGCAGTCCGAAAGGACGGAAAGCGGCGCG
TTCGAGCGTGAACAAGGGGTTGCCGCCGCCGTCGGTCAGGTCGAAACACTCGTTGCGCCA
GCCGTCCAACAGCCCCGCACAGTGCCAACCGAGGGCGAGGTGCTGTAAGCGTCCGCCCAT
ATCAGGCCAGCCGTCCGCATTCAGAAAAATGCCGTCTGAAGACTCCGAGCAGCCTGCCTC
CCAGTCTTTTTTGACGCGTTCCACCCATTCCGGCGACAGGTTGCCCAAAGGCAGACCGTT
CAGATACAGCGTTTTCCAGCAACTTTCTGCACCGTAACTTGCTTTTGCCCACTCGAACAG
AGCATCAAGGTCTTGTTTGCTGACGGATTCGGTAAAACGGACGGTCGGCATGGTTGGCTT
TCTTGAAGACAATGCGGAAGATTGTAGCCAAACTGCCTGAATTTGCAACCGCAAACAGAC
GGCGGTCTGCCGTTCCGATATTGCGGCCGGAAGGAAAATCCGGCAAAAAACGGAGGCGGC
AGGAAACAGGAAACAGGAAACAGGCAAAACAAAACCCCTCGGCATTGAACCGAGGGGTTT
GATATTTGGTTGCGGGAGCAGGATTCGAACCTACGACCTTCGGGTTATGAGCCCGACGAG
CTACCATGCTGCTCCATCCCGCGTCAGAAGATGAAACTATACGGCAGATTTTTTATGTTG
TCAAACATTACTTCCGCGAAAAATCATAAAATTTTGCCGTCCGGCGGTTTATGCCTGATT
TGAAATATTATTTTCTTTACAAAAGTTCATGTTCGTGATTTAATTTTGGTTAACATTGAA
ACAGGGGTGCTGCCTGATGTTTAGGCGGCTGAGAAATACCCTTTACACCCGATCGGGATA
ATACCTGCGTGGGGAGTTTTCACGGATTCTGCTTTTCAGACGGCATTGGTTTTCAAATGC
CGTCTGAAAACGCAAAACGCTCCTGTTTCTTTAATTCTAAACGAGAAAACAGGAGCATTT
TTTTATGACTACGCCAAAAAAAAACCGCCAAAACTTCCGGCAACGAAGCGCGCGAGCTTGC
CGACTTGAGCGAAGACATCGGCATCTGCTTTAAATATCCGAACTCGGAACGCGTGTATCT
GCAAGGCAGCCGCGACGACATCCGCGTGCCTTTGCGCGAAATCCGTCAGGACGACACCTA
CACGGCGCAAGGTACGGAAGCCAATCCGCCGATTCCCGTCTACGACACCAGCGGCGTGTA
CGGCGACCCGGCGGCGCATATCGACCTGAAACAAGGTCTGCCGCACATCCGCACCGCGTG
GCTGGACGAACGCGGCGATACCGAAATCCTGCCCAAGCTCTCCAGCGAATACGGCATCGA
ACGCGCACACGATCCGAAAACCGCCCATCTGCGTTTCAACCAAATCACCCGCCCGCGCCG
CGCGAAAAGCGGCAGCAACGTAACCCAGCTTCACTACGCGCGCCAAGGCATTATCACGCC
CGAAATGGAGTTTGTCGCCATACGCGAACGTTTAAAATTAGACGAATTGTCCCAAAAACC
GGAATACGCCAAACTCTTGGAACAGCACGCGGGGCAAAGTTTCGGTGCGAACATCCCGAC
CCATCCCGACCAAATCACGCCCGAATTCGTGCGCCAAGAAATCGCCGCCGGACGCGCGAT
TATTCCCGCCAACATCAACCACCCCGAACTCGAACCGATGATTATCGGCCGCAACTTTCG
TGTCAAAATCAACGGCAACTTGGGCAACTCCGCCGTCACCTCCAGCCTGACCGAAGAAGT
CGAAAAAATGGTGTGGTCGCTGCGTTGGGGCGCGGACACGATTATGGATTTGTCCACCGG
CGCGCACATCCATGAAACGCGCGAATGGATTATCCGCAACGCGCCCGTCCCCATCGGCAC
CGTGCCGATTTACCAAGCGTTGGAAAAAACCGGCGGCATCGCCGAAGATTTGACTTGGGA
TTTGTTCCGCGACACCCTCATCGAACAGGCGGAGCAAGGCGTGGACTATTTCACCATACA
CGCGGGCGTGTTGCTGCGTTATGTGCCGATGACCGCCAACCGCCTCACCGGCATCGTCTC
GCGCGGCGGTTCGATTATGGCGAAATGGTGTTTGGCACATCATCGGGAAAATTTCCTCTA
CACGCATTTCGACGAAATCTGCGAAATTATGAAAGCGTATGACGTATCGTTCAGCCTCGG
CGACGGCCTGCGCCCCGGCTGCATTGCCGATGCCAACGACGAATCCCAATTCGCCGAACT
GCACACCTTGGGCGAATTGACCGATAAAGCGTGGAAACATGACGTACAAGTCATGATCGA
AGGCCCCGGCCATGTGCCGCTGCAACGCGTCAAAGAAAACATGACCGAAGAGCTGCAACA
CTGCTTTGAAGCACCTTTTTACACGCTCGGCCCGCTCGTTACCGACATCGCACCCGGCTA
CGACCACATCACCTCGGGCATAGGCGCGGCCAATATCGGCTGGTACGGCACGGCGATGCT
TTGTTACGTTACCCCGAAAGAGCATTTGGGGCTGCCCGACAAAGAAGACGTGCGCACCGG
CATCATCACCTACAAACTCGCCGCCCACGCCGCCGATCTCGCCAAAGGCTGGCCGGGCGC
ACAATTACGTGACAACGCCCTGAGCAAAGCGCGTTTCGAGTTCCGCTGGCGCGACCAATT
TCGCTTAAGCCTCGACCCTGAACGTGCCGAGAGCTTCCACGACGAAACCCTGCCTGCCGA
AGGCGCGAAAATCGCCCACTTCTGCTCGATGTGCGGCCCCAAATTCTGCTCGATGAAAAT
CACGCAGGAAGTGCGCGACTACGCCGACAAGCAAAAAGCCCAACGGCAGGGCATGGAGGA
AAAAGCGGTCGAGTTCGTCAAAAAAGGCGCGAAGATTTACAGTTAAACGTCAAGCAAAAA
ATGCCGTCTGAAAACCGGAAAAAAGGCTTCAGACGGCATTCTTTCGCTTGTGAAAATATA

GTGGATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAGATAGTACGGAA
CCGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCA
ACGCCGTACTGGTTTTTGTTAATCCACTATAATTTTTAAAATTTTTATATTCACATAGAT
GGGGATGATGGGATTTAGGATTCTGATTTTGTTTTTGAGAGAATGAAGGAATTTGAGATT
GTGGGCGATTATCGGGAAAAATAGAATCTTTCCGCCGTCATTCCCGCGCAGGCGGGAATC
TAGACATTCAATGCTAAGGCAATTTATCGGGAATGACTGAAACTCAAAAAACTAGATTCC
CACTTTCGTGGGAATGACGGGATTAAAGTTTCAAAATTTATTCTAAATAGCTGAAACTCA
ACGCACTGGATTCCCGCCTGCGCGGGAATGACGAAGTGGAAGTTACCCGAAACTTAAAAC
AAGCGAAACCGAACGAACTAGATTCCCACTTTCGTGGGAATGACGGCAGAGCGGATTCTG
TTGCTCCCGATAAATGCCGCAACCTCAAATCCCGTCATTCCCGCGCAGGCGGGAATCTAG
GTCTGTCGGTGCGGAAACTTATCGGGTAAAACGGTTTCTTGAGATTTTGCGTCCTGGATT
CCCACTTTCGTGGGAATGACGGGATTAAAGTTTCAAAATTTATTCTAAATAGCTGAAACT
CAACGCACTAGATTCCCGCCTGCGCGGGAATGACGGCATATTTTGACATTGAATAAAAAA
GACTAAAAACAGGAAAAAGCCAAACAGAAAAAAGCCAAACAGAAAAAAGCCAAACAGAAA
ACAGCGAAACAGAAAACAGCGAAACAGAAAACAGCGAAACAGAAAACAGCGAAACAGAAA
ACAGCGAAACAGAAAACAGCGAAACAGAAAACAGGAAAAAGCCAAACAGAAAACAGGAAA
AAGCCAAACAGAAAAAAGCCTGTCTGGCGACAGGCTTTTTGTTGATACCAATCTTTGCAG
ATTAGAATTTGTGGCGCAGACCGACACCGCCGGCAGTCGCTACGAATTTGTTTTCGCCTT
TGCCTTCTTGCAACCAACCGGCAGAAACCAAGGCAGAAGTGCGTTTGGAGAAGTCGTATT
CCGCACCGACAACCACTTGGTCGTATTCGTTGCCTATGTCTGCATCATCAACCAAACCTT
TGAAGCCGTGGGCGTAAGAAACTCGGGGCGTTACGTTGCCGAAGCGGTATGCCAAGGTAG
CGGCAACTTCGGTTTGAGAGTTGTGCGAATTGGAAGCATCAGTCAGTTTCGCGTCTTGTT
GCTGTACGGCTACGGAAGCGTACAGGGCATCATTGTCGTAACCGCTGACCAAACGGTGAA
TCTGGTATTTCTCAATATTCAAGCCCTCTTGCACTTGATGATGTCTTTTATAGGCACCGC
CATATTGCACGAAGAAGCCACCGTTTTTGTAGTTGAAGCCGGCGTGGTAAGATTCGCTGT
TATGTCTGCCTGCATTGTCGTTAAGCGCGTATTGTACGCTGCCGCTGAGGCCGGCAAATT
CGGGAGAATCGTAGCGTACGGAAATGAGGCGTGCCTCGGGTTCGGCAATTTTGTTTACAC
CCAAATAGTCGCTTTTGCTATCCCAAGGATTGATGTCGCCGGTGTCTTTCAGGACGCTGT
TCAAACGACCGACGCGCAATTTACCGAAGCCGCCTTTCAAGCCGATGAAGGATTGGCGGT
TGCCCCAACCGGAGTCAGTACCGGCGATAGATGCTTTTTGCTCAACCTGCCAAATGGCTT
TCAGGCCGTTACCGAGGTCTTCTTGGCCTTTGAAGCCGATTTTCGAACCCAAATCAACGA
TGCCGGTAGCGGTTGTAACTTCAGTAACTTGGCCGTTCTGGTGAAATACAGAGCGGGAAG
TTTCTACGCCGGCTTTGATGGTGCCGTACAGGGTAACGTCAGCCATTGCTGCAACAGGAA
GGGCTGCCAAAGTCAGGGCAATCAGGGATTTTTTCATTGCTGTATTCCTTTTTTGGTTAA
GAAATTTAAGCCGGCCGGGCTTTCCAAGCCGCTTAGCTTTGCATTTACCGCCGACGTTTT
CGGCGATTCCATGTCTGAAACTATAGATGTTCCTGTCTGAAAAAACAAACTTCTGCCTGT
TTTTGTGTGGAATTCGATGTGTATTTTGAAGGGCGGATAGAGATATTATGGGTATTTTTT
GTTTTATAACATATGGTTATTTAAATTTTTTAAGATTTGCATTTTTACAACACTTACTCG
GGAGGGTATTGGAGGGCATTGCAAACCGGGGGTTATAAAGACGGCAAAAAACGCGCACGT
GGTTTCTTTCGGAACGGGCGGATGCAAAACCCGCCCCTGCGGCAGGAGATAGTGGATTAA
CAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAAATAGTACGGAACCGATTCAC
TTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCTAACC
GTCATTCCCGCCACTTTTCGTCATTCCCGCTCAGGCGGGAATCTAGAATCTCGGACTTTC
AGATAATCTTTGAATATTGCCGCTGCCTTAAGGTCTGGATTCCCGCCTGCGCGGGAATGA
CGGCTGCAGATGCCCGACGGTCTTTAGAGTGGATTAACAAAAATCAGGACAAGGCGACGA
AGCCGCAGACAGTACAAATAGTACGGAACCGATTCACTTGGTGCTTCAACACCTTAGAGA
ATCGTTCTCTTTGAGCTAAGGCGAGGCAACGCTAACCGTCATTCCCGCCACTTTTCGTCA
TTCCCGCTCAGGCGGGAATCTAGAATCTCGGACTTTCAGATAATCTTTGAATATTGCTGT
TGTTCTAAGGTCTAGATTCCCGCCTGCGCGGGAATGACGAATCCATCCGCACGGAAACCT
ATATCCCGTCATTCCTACGAACCTACATCCCGTCATTCCCTCAAAAACAGAAAACCAAAA
TTAGAAACCTAAAATCCCGTCATTCCCGCGCAGGCGGGAATCCAAACTTGTCCGCACGGA
AACTTATCGGATAAAACGGTTTCTTAGATTCCACGTTCTAGATTCCCGCCTGCGCGGGAA
TGACGAATCCATCCGCACGGAAACCTATATCCCGTCATTCTTACGAACCTACATCCCGTC
ATTCCCTCAAAAACAGAAAACCAAAATTAGAAACCTAAAATCCCGTCATTCCCGCGCAGG
CGGGAATCTAGGTCTGTCGGTGCGGAAACTTATCGGGTAAAACGGTTTCTTTAGATTCCA
CGTTCTAGATTCCCGCCTGCGCGGGAATGACGGCTGCAGATGCCCGACGGTCTTTATAGT
GGATTAACAAAAATCAGGATAAGGCGACGAAGCCGCAGACAGTACAAATAGTACGGAACC

GATTCACTTGGTGCTTCAACACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCGAGGCAAC
ACCGTACTGGTTTAAAGTTAATCCACTATAATGAACACAATCCATTCAGACTATTCAATC
AGGCAAACATCTCCTGCAATACTGCAAACAGTTTTTCAGCCGTACTGTTGTCTAAATTGC
CAAGATGTTTGACCAATCCGGCTTTATCCACAGCCCTAATCTGTTCGGGCAAAAGCAAAC
CGTCTTTATCCTGAAAGCGGACATTGACGCGGAACGGGGCAGGACGGCTTCCGCTCGTCA
TGGGAACGATCAGCACAGTCTTGAGATAGTTGTGTATTTCAGGAGGAGAGACTACGACAC
AAGGACGTGTCTTTTTGATTTCGCTTCCTACGGTCGGGTCTAAGGAGACCAGATAGATTC
CGCCGCGTACTACCATATCCATTCTTTATCCGCTTCGTTTTCAATTTCGGAAAAAAAATG
CTCCTGCTCGGTTTCGACAAGCATTGCGGCAGCTTCTGCCCATCCCCTGCGAACGGTAGG
ACAGCTTAAAATAATATTGCCCTTTTCAACTGTAACAGCCAAGCTGTCTACTGCCCCTAT
TTGACCCAATAATGATTTGGGCAGAATCACGCCTTGCGAGTTTCCCATTTTGCGTATGTT
GAGAATCATATACGTACCATATTCACCTGTTTATGTAATAACAATGTTAGTACCTTGATG
AGGTAGTGTCAACATGGAAAAAGATTGCCTGACAGTTTGTCCGATTTCAAAATCTCCGCG
ACAAGCATGTTTTAAAGCCATTCGGGGATTTGGGGGCGGATGATGCCGTATGCCTCGGGA
TAGTCGACGCCGGTCAGGTAAAGTCCGTCGGGCATGAAGGTCGGCGGGGCTTTGAGGCGG
CTGCGTTCTTGAATCAGTGCGGCGAAGCCTTCGACGCTGAGTCTGCCGCTGCCGACATAA
ACGAGCGCGCCCATGATGTTGCGTACCATGTGGTGCAAAmAgGCGTTGCCGTGCAAATCG
AGGCGGACGAGTCCTGAGCTTTGGGTAAGGTTCGGCGCGGTAGATGGTTTTGACGGGGGA
TTTTGCTTGGCATTCGGCGGCGCGGAAGCTGGAGAAGTCTTGTTCGCCGACCAATAAGGC
GGCAGCCTGCCGCATCTGCCCGATGTCGAGTTTGAGGTGTGTCCAGCCTGCCCTGTTTTT
GAGCAGGGGGGAACGGACGGGGGCGGATTCGAGCAGGTAGCGGTAGTGCCGTCCGTATGC
GTCAAATCGTGCATGAAATTCGGGGGCGACCTGTCGGGCGTGCAAAACGGCAATGCCTTC
GGGCAGGTGGGCATTTACGCCGCGCACCCATGCCTGTTGGGGACGGGCGGCAGTTGTGTC
GAAGTGGACGACTTGGGCGGTGGCATGCACGCCGGTGTCGGTCCTGCCGGCAACGGTGGT
GGAAACCGCTTCCCCTGCTATTTGGGCGAGCGCGGTTTCCAATGCCGCCTGAACGGTCGG
TACGCCGTCAGCCTGTTTCTGCCAGCCGTAAAAGCGGCTGCCGTCATAGGATAGGGTTAT
TGCCCAGCGTTGTTTTTGTGCGGTATCCATCGGATTTGGGATTCGGATAAATGTTCAGAC
TGCATTGTATCGCAGATTTTGCAGGGAAACGGCAAACGCCCAGGGCGAGCGGCGTTGTTT
GGGGAGTTGTTGGGGGGGGGGGATGCAGTTGCTACGAATCGCTATCCTGTGAATTTACCC
TGTCAGGAGTGCCCGAATCGTCATTCCCGCGCAGGCGGGAATCTAGGACGTAAAATCTAA
AGAAACCATTTTATAGTGGATTAACAAAAACCAGTACAGCGTTGCCTCGCTTTAGCTCAA
AGAGAACGATTCTCTCAGGTGCTGAAGCACCAAGTGAATCGGTTCCGTACTATTTGTACT
GTCTGCGGCTTCGTCGCCTTGTCCTGATTTTTGTTAATCCACTATATCCGATAAGTTTCC
GCACCGACAAAACTAGATTCCCACTTTCGTGGGAATGACGGGATGCAGGTTCGTGGGAAT
GACGCGAACAGAAACCTCAAATCCCGTCATTCCCGCGCAGGCGGGAATCTAGACCTTAGA
ACAACAGCAATATTCAAAGATTATCTGAAAGTCTGGGATTCTGGATTCCCACTTTCGTGG
GAATGACGGAATGTAGGTTCGTAGGAATGACGGGATGCAGGTTTCCGTATGGATGGATTC
GTCATTCCCGCGCAGGCGGGAATCTAGGTCTGTCAGTGCGGAAACTTATCGGATAAAACG
GTTTCTGGAGATTTTTCGTCCTGGATTCCCACTTTCGTGGGAATGACGCGGTGCAGGTTT
CCGTATGGATGGATTCGTCATTCCCGCGCAGTCGGGAATCTAGACATTCAATGCTAAGGC
AATTTATCGGAAATGACTGAAACTCAAAAAACTGGATTCCCACTTTCGTGGGAATGACGG
GATGCAGGTTTTCTTAACCCCGCGTTCTAGATTCCCACTTTCGTAGGAATGACGGCGGTA
GATTTGGCAGATGCGGCGGATTCGGCAGGTCTCAACCCATCCTACAATCCACCCTGACCG
CCTTTTACGAATCCGCCGCCATCCTGGGAAACGCAAAAAAAATGCCGTCCGAAAACCTTT
CAGACGGCATTTTCGCGGGCAAATCAGTAAAAGACTTCGCGTCAGCTTAAGCGTTTCATA
CCGCACGTCGGACCGGGCGGGTTTCGGGTTTCGGAAGAGCGGTGTGAAATGAAACGCGGC
AGACGGCGTGTGTGCCGCCATCCTGCCCCGAACCGGCGGCAAAGCGTCATCGCGGTTTGA
ATCGGGCGGGCGCGTGCCTGTTCCGGCACGGATGCCGTCTGAAAGCGGCGGGTCGGTCCG
GTCCGCGTCAGCCTAAGGCGAAAAGTTCCCTGCCGTGGATTTTGAGCCACTGTTTCGCTT
TGGGCGTGTAGGGGGCGAAGCGGCGGGTCAGTTCCCAAAAGGCGGGGCTGTGGTCGGGAT
GGGCGAGGTGGCAGAGTTCGTGTATGCAGACATAGTCGGCAACGTATTCCGGTGCGCCGA
CCAGCCGCCAGTTGAAGCGTATGCCTGTGGTTTTGCGGCACACGCCCCAGAAGGTTTTGG
CAGAGGTCAGCGAGGAGGAGGCGGGGAACAGTTGTGTGGTGCGGGCGTGGCGTTCGAGGC
GGGGAATCAGGTAACTGTGCGCCTGCCGTTCCAAAAAGTCCCGCAGCAGCGCAAGCTGTT
TTTCGGGTGCGCCTTCGGGAACACGGATTTCAGACGGCATCAGCAGGATTTGCGTGTCTT
GATGGGCGGTGAGGGCAAGCTGTCTGCCGTGGAAGAGGATGGATTCGGGCAGCCGGTTTT
CGGCAGTTTGCGGCGGCGGTGTTTTCGCCAGTGTTTGCCGCAGGACGGCTTCGTTTTCAT

ACAGCCAGCGGTTTAGAGCGGAGACGGAGAAGCAGGGTGGGACGCTGATGCGGACGGTAT
GTGTGCCGGCGGGGCGGATAATCAGGTTTTTCTTGGCACGGCGCTTGATTTCGACGGTCA
GTTCCATGCCGTCTGAAAGGGTGTGGACAAAGGCGGTCATGCGGTTTCAGACGGCATTTT
GGCGGCAAACGGGCCTGCGCCGGAAATAAGCGGTTGTTGCGTTTCGATGAGATGTTCGCA
TTTTTCCATCAATTCGGCTTCGCTGCCGCTTGCGTGCGGGATGGTCGGACAGATGACGAC
GGTGATTTCCCCCGGATATTTCAGAAAGGAGTTTTTCGGCCAAAATTCGCCGCTGTTGAG
GGCGACGGGGACGATGTCCATCTCAAACATTTTCGCCATGCGCGCGCCCGCCGAGTTTGTA
TTTGCCGCGTTTTCCGGGCGCAAGGCGCGTGCCTTCGGGGAAAATGGTAATCCAATAGCC
TTCGTTTTTGCGCACCAACCCCTGTTTTATGAGCTGCTCGTTGGCTTCGCGGCGGTTGTT
GCGGTCTATGCCTATGGTTTTGACCAGTTTCAAGCCCCAGCCGAAAAAGGGGATTTTGAA
CAACTCGCGTTTGGCAACGTAAACCTGCGGCGGAAAAATGTCCTGAAGGGCGAGCGTTTC
CCAGCCGCTTTGGTGTTTGGCGCAGATGACGGCGGGGCGGTCGGGGATGTTTTCCGCGCC
GATGATGCGGTATTTGAGCCCGACGATGTGTTTGAGCGACCAGTTGAGAATGCCGACCCA
GACCCGCGCCATCTTGTGCGCCCCGTCCCGGAAAGGCGAGGCGAGCAGCATAAAGGGAAA
GAGGAAAATCAGGGTGGAACAGAGTATCAGCCAGTAAATCAGGTTGCGGATGATGAGCAT
GGTTTTGCCTTGTCGGAATGCGGTATGTTCAGTCGGCTTGCGGTGCGGTGTTTTCCTGCA
TGATGTATTGTGAGAAATCGAGCAGGGTATCGAAAACCTGTGTGTGTTCGGGCAATTCGT
GTCCGTGTTGGGAGAGCGTTTTTTTGCCTTTTCCGGTCAGAACCAGCGCGGGTTTGCCGC
CGACGGCATCGATTGCCTGCAAATCGCGCAGGCTGTCGCCGACCAGCCAGGTTTCCGAGG
CTTGGGCGTTGAAGCGTCCGATGATGTCTTCAATCATACCCGGTTTGGGCTTGCGGCAGT
TGCAGTTGTCGGCATCGGTGTGCGGGCAGAACCAGATGCCGTTGATTTCGCCGCCTGCCT
GACGGACGAGGCGGTGCATTTTGGCGTGCATTTCGGTAAGGTTTTGAACGGTAAAATATT
TGCGCCCGATGCCGGATTGGTTGGTGGCAACGGCGACGGTGTAGCCTGCCTGCGTCAGAA
ATGCCACCGCATCCATGCTGCCTTCGACAGGTATCCACTCGTCAACGGATTTGACGAAGT
CGTCGCGGTCCTGATTGATGACGCCGTCGCGGTCGAGAATGATGAGTTTCATCGCGTTTC
CTTTGGATTGGGGCAGGTCGGGGGTGGCATTATACTGAAATATCGGTGGAAATGCGCCTG
TGCCGCGCGATAACGCGCCTGTTCCGGCAACCGCTTGCAATGCGCGGGTACGGCGTTTCG
GGGGCGGTTGCCGATGTTTTCCTGCCTGCTCCTGCAACGGCGGATTTTCTGCCGGACGTA
TTTTCCGCTGCAAAAGTCTCGGCGCGGAAGTTGCGGGGCAATGCCGTTTGGGGACGCGGT
CAGGCGGCAAGCATTTCAAGAATCAGCCCGATGTGTTTGTCTAAAGCCGGGTTGTAGTTC
AGTGCCCGCATCGGGTCGGACAGGCGGTGTCCGCCCTGTTTTTAAGGCTGACGGCTTGG
GCAACGGCTTGCGCCAGGGTTTCCGGGTTTTGGCGTGAGAAAAAGAAGCCGGCTTCTTCG
TTCATGACCTCTGTACATGCCATGTTTTCGGAGAGGACGACGCGTGTGCCGCATAGGACG
GATTCGACGCCGACCAGCCCGAAGGGTTCGTACAGGGAAGCCATAATGGTAAAGTCGGCG
GCGCGGTAGAGTTCGGGCATATCGGTGCAGAAGCCCAGTCCGACGACGTTTTTCATAGGG
CGGGGAAGCGGGGAGCCGACAACGGCGAGCTTGACGGGCAGGCTGGTATGTTCGAAAAAG
TCGGCAAGCAGTTCCAGACCTTTGCGCGTGTGGCCGGTCGATGGGAACAGGAAAACGGTT
TCATGGTCGGCAAAGCCGTATTTGGCGCGCAGGTCGGCAGTTTCTCCGGGTTGTGGAAAG
AAGCGTTCCGTATCTGCGGGGGGGGGGGGCGACTTGGATTCTTTCAGGGGGAACGCCGTA
CAGTCCGACCAGTTCGCGCCGCATCATATGGGAATGCGCCATAATCAGTTTGGCGGTGGC
GTAGTTGCTGCGGTTGCGGCGTATGGCGAGGCGGTCGAGCAGGTTCGGTTTTTGCGCCAT
ATGGTGCAGGTAGCCCAAGTGTGTGCCGCCGCAGATGAGGAGGTCGGCGTAATCGGCGTG
GTGGCAGGCAATCAGTTTGGCGGCACTGTTTTTTCTGGTTTGAGCGAGCCGGCTTGAAAA
GAGGAATGAGCGTAGTTTTTTCAGCGTCCGGTGTTGATCGACAAGATGGGGTTCGATTAG
GGCGTATTCAGGAATGCTGTGATCAAATTTCGTCGCATAAACGGCCGGTGTGATGTTTTG
TCTGTTCAGACCCTTTACCAAATCCAATGTGTAGCGTTCCGTGCCGCCGCCGTGTTTGAA
GTTGTTGGTTGCAATGTCTATTTTGAGCTTCATCATTGTTCCTTTATGGTTGCGTCCCGG
TTTGTCGGGGCGGGATTTGTGCGTGAGGGGCAGGGTAATGCGCTGTGTGCCGGAATACGG
TTGCCGTTTGTTGCGGCAATGCCGTCTGAAGCCGCCGGCGGGCTTCAGACGGCATTTTGC
CTTTATCCTTTAAATACGGGGACGAGTTCCATTTGGCTCAATACCTGTGCGGCGATGTTG
ACGATGCCGAAAAGGAAGACCCAAACCATCAGCCACAAGCCGCCGTAAACTTTATAGGTT
TTGCCTGCGCCGAATTTTTTGCGCGAACGGTAGAGCAGCATGGCGGGGATGATGCCTGTC
CAGACGGTTGCCGCCAGGCCGACGTAGCCGATGGCGGTAACGAAGCCGGTGGGGAAGAGC
AGGCAGGAAATCAGGGGCGGCAGGAAGGTCAGCGCGGCGGTTTTGGTGCGGCCGGAGATG
CTGTCGTTCCATTTGAAGATGTCGGCGATGTAGTCGAAGAGTCCGAGCGTTACGCCTAAA
AACGAGGTGGCGATCGCCATATAGGAAAACAGGGACAATATTTTGTCCATATTGCCGGTT
TGGGCGAATTTGGACAGGGTTTCGATGAGGACGGAGACTTGCCCTTCGGCGGCGATGACG

GGGGCGAACTCGTTGCGCGGCAGGTTGCCTTGGATGGCGGTTTGCCAGAGGACGTAAATT
ACCAGCGCAATCAGTGTGCCCGTCCAGATGGATTTAGCCACTTTGGGCGCGTCGCCTTTA
AAGTATTTGAGCAGGCTGGAGACGTTGCCGTGGAAGCCGAAGGAAGCGAGGCAGACGGGC
AGGGCGGTGGCGGCGTAAATCCAGTAGTTTGTGCCGGCGGGGGCTTGGGTATCGAAGAGG
ACGGACGGCTTGGCATCGGCAATCAGCCCGCCGGCCGCCCAAATAAAGGTCAATACCATG
CCGCCGATAAGGACGCCGGTGAAGCGGTCGACCAAGCGTGCGGATGCCCATACGCAAAAG
GCGAGGATGCCGAAGAAGACGAGTTGTCCGACGGTGAGTGAAACGTCGCCGCCTGCCGCG
CTGCCTAAGCCTTTGGCGGTCAGGTCGCCGCCGACGAAGATATAAGCGTAAGTAAGCAGG
TATAAAACGAAGGCGACGGCGATGCCGTTGATGATGTTCCAGCCGCGTCCGAGCAGGTCT
TTGACCATCGTGTCGAAACTTGCGCCGTGCGGATAATGGGTGTTGACTTCCAAAATCATC
AGGCCGCTGGAAAGCATAGAAAACCAGGTGTACAGCAACACGGCCAGCGAGCCGGTAAAC
CATACGCCGGATGTGGCGGTCGGGTTGGCGAGCATGCCTGCGCCGATGACCGTGCCGGCG
ATAATCATCGCGCCGCCGAACAGTGAAGGGGTTTTGTTGGGCATATTTTGTCTTTCTGCC
AGAAAAAGCGAGCCGCCATTATGCCGTAAAGTGTAAGGATTTGTAAGGTATTTGCGCCGC
GCCGCCCGAAAAGGCTTTCAGACGGCATTGTGTTCCATAGTATAATCTTGGGTTTTGGAG
TGGGCGGTTCGTCAGATGGGAGGGAAAATGTCCGACAAAAAATATAATGTCGATGAGGGG
GAAATCGCCAAATTCAGCCGGATTGCCGACAAATGGTGGGACAAGTCGGGCGAGTTCAAA
ACCTTGCACGACATCAATCCGCTGCGGCTGGATTATATCGACGGACACGCGGATTTGTGC
GGCAAACGGGTTTTGGACGTGGGCTGCGGCGGCGGCATCTTGGCGGAAAGTATGGCGCGG
CGCGGCGCGGCGTTTGTAAAGGGCATCGACATGGCGGAGCAGTCGTTGGAAACCGCCCGC
CTGCACGCGGCTTTGAACAATGTCGCCGATATCGAATACGAATGTATCCGCGTGGAAGAC
CTTGCCGAGGCGGAACCGCACTCGTTCGATGTGGTAACGTGCATGGAAATGATGGAACAC
GTCCCCGATCCCGCCGCCATCGTGCGTGCTTGTGCCAATCTGGTCAAACCGGACGGCATG
GTGTTTTTTTCCACCATCAATAAAAACCCGAAATCGTACCTGCATCTGATTGTGGCGGCG
GAATATCTGTTGAAGTTTGTCCCCAAAGGCACGCACGACTGGAAAAAATTCATCGCACCT
GCCGAGCTGGCGCGAATGTGCCGTCAGGCAGGCTTGGATGTGGCGGATACGAAGGGTATG
ACTTACCATGTGTTGTCGCAAACTTATGCCCTGTGCGATTCGACCGATGTGAATTATATG
TTTGCCTGCCGTCCGGCGTTCTGACGGCGGGTTTGCCCGTTTTTGAGCAAGTGAGTTGAT
ATGTCTGTCTATACCAGTGTTTCCGATGATGAAATGCGCGGCTTCCTGAGCGGTTACGAT
TTGGGGGAATTTGTTTCCCTGCAGGGCATCGCGCAGGGGATTACCAACAGCAATTATTTT
CTGACGACGACTTCGGGACGTTATGTGCTGACCGTGTTTGAAGTGTTGAAACAGGAAGAG
CTGCCGTTTTTTCTGGAGCTTAACCGGCATTTGAGTATGAAGGGCGTGGCGGTTGCCGCG
CCGGTTGCGCGCAAAGACGGCCGGCTTGATTCCGTTTTGGCGGGTAAGCCTGCCTGCCTG
GTTGCCTGCCTGAAAGGTTCGGATACCGCGCTGCCGACGGCTGAGCAGTGTTTTCATACC
GGTGCGATGTTGGCGAAAATGCACCTTGCCGCCGCCGATTTCCCTTTGGAAATGGAAAAC
CCGCGTTACAATGCGTGGTGGACGGAGGCGTGCGCCCGGCTGCTGCCCGTCCTGTCGCAA
GACGATGCCGCACTGCTGTGTTCCGAAATCGATGCGTTGAAGGACAATCTCGGCAATCAT
CTGCCTTCGGGCATCATCCATGCCGATCTGTTTAAAGACAATGTGTTGCTTGACGGCGGT
CAGGTATCGGGCTTCATCGATTTCTATTACGCCTGCCGGGGCAATTTTATGTATGACTTG
GCGATTGCGGTCAACGATTGGGCAAGGACGGCGGACAATAAGTTGGATGAGGCGTTGAAA
AAGGCGTTTATCGGCGGTTATGAGGGCGTGCGCCCCTTGAGTGCCGAAGAAAAGGCGTAT
TTCCCGACCGCCCAACGTGCCGGCTGCATCCGTTTTTGGGTGTCGCGCCTGTTGGATTTT
CATTTTCCGCAGGCGGGCGAGATGACGTTTATCAAAGACCCGAACGCGTTCCGCAACCTG
CTGTTGAGTTTGGGTTGAGTGCGTCCGGCGTTTGACAGAAATGCCGTCTGAAAGGGTTTC
AGACGGCATTTTTATGGCTGATTAAAACGAAAATGAGACGATAGCCGGGTATTTTCCATT
TTATAGTGGATTAACAAAAATCAGGACAAGGCGACGAAGCCGCAGACAGTACAGATAGTA
CGGAACCGATTCACTTGGTGCTTCAGCACCTTAGAGAATCGTTCTCTTTGAGCTAAGGCG
AGGCAACGCTGTACTGGTTTTTGTTAATCCACTATAAATCAATGTAATCCATTCTGTTCC
CGATGTTTATGCCGTCTGAAACCCATCCTGTGTCGGGCTTCAGACGGCATATTGCTTCAA
AGCAGGTTTTCCGAGGCAACCCAGTTCAGAATATCGGCTTCGACCGCTTCGGGCGTGCCG
GGGTTGGCGATTTTGACCCCGTATTCGCCTTCGCCCAGTTCCTCGAGGATTTCCAGTTGC
GAATCGAGCATCCCTGCTTTCATGTAATGTCCTTTGCGCGACATCATGCGCTCGAGGTTG
ATGTCTTGCGGCGGACTGAGGTGGATGAAGGCAGCTTTGCCTTCGGCTCCGCGCAGAATG
TCGCGGTAGCCGCGTTTGAGGGCGGAACAGGTTACGATGGTGTGGTTCGCACCGTTTTGC
GCCTGTTGCGTCATCCAGTCGCGCAGATTGCCCAACCACGGATAGCGGTCTTCATCGGTC
AGCGGAATACCCGCGCCCATCTTGTCGCGGTTGGCTTGGGTGTGGAACTCGTCGCCTTCG
GCATAGGGACATTGACCGAGGTGTTTCTGCAGGGACAGCGCGGCGGTGGTCTTGCCGCAG